The Price Is Right: Examining Demand for Medical Care in the Presence of Deductibles

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Helen Markelova Trenz

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Roger Feldman and John Nyman

December, 2015

Acknowledgements

This five-year long journey was instrumental in my life, both personally and professionally. I will forever be grateful to the support and encouragement from my professors, friends, and family that I received along the way.

First and foremost, I owe my deepest gratitude to my advisors, Roger Feldman and John Nyman. I fell in love with the field of health economics in John's class during my first semester at the University of Minnesota and learned so much about this field as his research and teaching assistant. I first came upon my dissertation topic during Roger's course, and I am so grateful for his encouragement to pursue this topic even when the research process faced challenges. John and Roger are my professional role models.

I am also indebted to Paul Glewwe, my committee member, for encouraging me to apply to the Department of Applied Economics, for sparking a love for microeconomics and econometrics through his courses, and for providing valuable input into my professional development and dissertation over the years. Pinar Karaca-Mandic was invaluable in making this dissertation happen by facilitating access to the MarketScan data and giving helpful comments during the dissertation. I have greatly benefitted from the discussions about my dissertation with Katy Katy Kozhimannil and Jean Abraham. I am grateful for the valuable input provided by Tim Beatty, Elizabeth Davis, and Gerard McCullough during my tenure at the Department of Applied Economics. My conversations with them helped shaped my professional development and career path.

I would also like to acknowledge the financial support I received from the Department of Applied Economics, Division of Health Policy and Management in the School of Public Health, the Graduate School, and the Healthy Foods Healthy Lives Institute of the University of Minnesota. This financial support has been invaluable in providing me the opportunities to pursue teaching and research opportunities and write my dissertation.

My community of family and friends was an unwavering source of support over the last five years. Martha Rogers and Brooke Krause were my "army buddies" over the last five years and helped me get through the coursework and dissertation writing. Diane Muchow, Erika Crowl, and Erica Danielson were my constant pillars of moral support and helped me to keep going. I am so grateful to my parents, Valerii Markelov and Galina Markelova, for always supporting all my endeavors and encouraging me to pursue my dream of getting a PhD. My grandmother, Valentina Popova, was my biggest cheerleader and motivator– she always believed that I could do it. My sons, Ashton and Weston, were my source of inspiration and joy. And, last but not least, I am so blessed and thankful to my husband, Matt Trenz. Words cannot express how grateful I am for your support, patience, listening ear, care, and encouragement during the dissertation process.

Dedication

To my three guys, Matthew, Ashton, and Weston.
You made it all worth it.

Abstract

Research in health economics has traditionally considered only the current price of care in the estimations of demand for medical services. However, given the typical structure of insurance contracts that include cost-sharing features such as deductibles, the price of medical care is not constant throughout the year and depends on past and future medical expenditures. This study explicitly incorporates this nonlinearity by using the more appropriate concept of expected end-of-year price and applying it to the analysis of the demand for medical care by a sample of insured pregnant women who face different end-of-year prices depending on the timing of labor. Additionally, it investigates whether this group of consumers is myopic or forward-looking by examining which price, current or expected end-of-year, women use when making purchasing decisions. The results show that women who give birth in a calendar year face lower expected end-of-year prices, but combined with other health factors, use less non-pregnancy related medical care than those who do not give birth within the same period. The findings point to the presence of forward-looking behavior, while not fully rejecting myopia. Additionally, when the probability of reaching the deductible is used as the price-changing event, rather than labor, there is more evidence of forward-looking behavior among women in the sample, as those who reach the deductible spend more on medical care in response to the lower end-of-year price.

Contents

Acknowledgements **i**

Dedication **iii**

Abstract **iv**

List of Tables **vii**

List of Figures **ix**

1 Introduction **1**

2 Literature Review **8**

3 Conceptual Framework **16**

4 Data and Sample Selection **21**

4.1 Data Description and Sample Construction 21

4.2 Selection Issues 23

4.3 Descriptive Statistics 27

5 Empirical Approach **32**

5.1 Main Idea 32

5.2 Modeling Issues 34

5.3 Empirical Methods 39

5.3.1 Least Squares Methods 39

5.3.2 Generalized Linear Model 42

5.3.3 Extended Generalized Linear Model 43
5.3.4 Two-Part Models 44
5.3.5 Count Data Models 47

6 Labor as Price Effect in Demand Analyses 50
6.1 Introduction and Motivation 50
6.2 Empirical Approach I: Labor as a Proxy for Expected-End-of-Year Price 51
6.3 Empirical Approach II: Labor as Instrument for Expected End-of-Year Price 55
6.4 Empirical Approach III: Month of Labor as Proxy for Anticipatory Effects 66
6.5 Summary 70
6.6 Sensitivity Analyses: April-September Births 72
6.7 Discussion 75

7 EFP in the Demand Equation 79
7.1 Introduction and Motivation 79
7.2 Empirical Set-up 80
7.3 Results 83
7.4 Sensitivity Analyses: Plans with Higher Deductibles 87
7.5 Discussion 91

8 Deductibles and Trimester Demand 94
8.1 Introduction and Motivation 94
8.2 Sample Construction 96
8.3 Empirical Set-up 100
8.4 Results: Labor as Proxy for EFP 101
8.5 Results: EFP in Demand Equation 110
8.6 Discussion 113

9 Conclusions and Implications for Policy and Research 115

References 122

List of Tables

4.1 Summary of Selection Analysis 25
4.2 Demographics and Healthcare Use 30
4.3 Characteristics of Insurance Plans 31
5.1 Model Fit Tests 38
6.1 Labor as EFP Proxy Estimation Results: Expenditures 53
6.2 Labor as EFP Proxy Marginal Effects: Expenditures 55
6.3 Labor as EFP Proxy Estimation Results: Utilization 56
6.4 Labor as EFP Proxy Marginal Effects: Utilization 57
6.5 First Stage Equation: Determinants of EFP 60
6.6 Labor as Instrument for EFP: Expenditures 64
6.7 Labor as Instrument for EFP: Utilization 65
6.8 Price Elasticities 66
6.9 Month of Labor: Expenditures 69
6.10 Month of Labor: Expenditure Marginal Effects 70
6.11 Month of Labor: Utilization 71
6.12 Month of Labor: Utilization Marginal Effects 72
6.13 Sensitivity Analyses: Descriptive Statistics 74
6.14 Sensitivity Analyses Results: Expenditures 76
6.15 Sensitivity Analyses: Marginal Effects 77
7.1 Probability of Reaching the Deductible 84
7.2 Probability of Reaching Deductible and EFP 85
7.3 Annual Demand: Expenditures 86
7.4 Annual Demand: Utilization 88
7.5 Probability of Reaching Deductible and EFP: "High Deductible" Sample 89

7.6 Demand Estimations: Sensitivity Analyses 90
7.7 Marginal Effects and Elasticities 91
8.1 Trimester Sample Construction 97
8.2 Labor and Non-Labor Groups by Pregnancy Cohort 98
8.3 Descriptive Statistics: Trimester Sample 99
8.4 Trimester Expenditures 102
8.5 Marginal Effects: Trimester Spending 103
8.6 First Trimester Expenditures 104
8.7 Marginal Effects: First Trimester 105
8.8 Second Trimester Expenditures 106
8.9 Marginal Effects: Second Trimester 107
8.10 Third Trimester Expenditures 108
8.11 Marginal Effects: Third Trimester 109
8.12 Probability of Reaching Deductible: Trimester Sample 111
8.13 Probability of Reaching Deductible and EFP: Trimester Sample 112
8.14 Expenditures during Pregnancy 112
8.15 Marginal Effects and Elasticities 113

List of Figures

3.1 Nonlinear Budget Set 17
5.1 Density for Non-Pregnancy Expenditures 36
5.2 Density for Log of Non-Pregnancy Expenditures 36
5.3 Density for OLS Studentized Residuals, log scale 36
6.1 Distribution of Expected Marginal Price 58
6.2 Labor Month Frequency Distribution 68

Chapter 1

Introduction

The ongoing discussions and gradual implementation of the new healthcare law, the Patient Protection and Affordable Care Act (ACA), which will require most individuals in the U.S. to carry health insurance, have fueled the long-standing debates among researchers from different disciplines, including economists, about the optimal design of health insurance contracts. With the increase in the number of the insured through the insurance exchanges and the expansion of Medicaid, the questions of how to tailor insurance contracts for both the new and existing policyholders are becoming relevant again, especially in the context of rising healthcare costs and the attempts to contain this growth. The key question is how to design policies that balance the tradeoff between moral hazard and risk protection, and the answer depends, largely, on the definition of price for medical care used in determining the value of health insurance.

To this day, the majority of this research in economics has only considered a single price of medical care that does not vary throughout the year. The results of these studies have been used to make influential conclusions about the individual and societal value of health insurance as well as to make recommendations for the design of health insurance policies. However, a typical insurance policy in the U.S. usually includes cost-sharing arrangements that make the patient responsible for part of the expenses. These features include a deductible, a co-insurance rate[1] that is usually applied to the expenditures

[1] Many health insurance policies include copayments, which is a fixed dollar amount, instead of coinsurance, which is a fixed percentage of expenses. This paper considers the co-insurance rate in the model. It should be noted that the presence of copayments instead of co-insurance also leads to a multitiered pricing structure that is discussed in this paper, even though the nature of the nonlinearity

after the deductible is met, and a stop-loss, which is an out-of-pocket maximum that a consumer would pay. Moreover, insurance policies are renewed for another year or selected as a new policy at the beginning of the enrollment period, which is usually January 1 of every year, so the spending thresholds are reset every year.

Hence, each year a person who enrolls (or re-enrolls) in a plan with a combination of a deductible, co-insurance and stop-loss does not face the same price of medical care throughout the year. Given the amount spent since the beginning of the calendar year prior to a new episode of illness, the consumer may have met and exceeded the deductible, which means that the out-of-pocket payment (i.e. the price she pays for care) for this illness would be lower than the full price she paid before reaching the deductible. In other words, within each calendar year, an individual with a typical health insurance policy faces a 3-tiered pricing structure that depends on the expenditure thresholds (amounts of deductible and stop-loss) stipulated by the policy. This nonlinear budget constraint may induce this policyholder to consume a different amount of medical care than she would have under a single-price policy (for example, under a policy with no deductible and same co-insurance rate for every medical procedure). At the same time, decisions to purchase medical care today affect the price of medical care tomorrow under a policy that includes nonlinear cost-sharing elements. Moreover, these decisions happen in the context of uncertainty because a consumer is likely unaware of her demand for care in the future.

Thus, the findings of the majority of the studies that rely only on the single-price definition of the price of medical care may not correctly describe how consumers demand medical care when they are insured and how they respond to the changes in the price of care. Some of these studies look at policies with multi-part pricing, i.e. cost-sharing elements, but analyze consumers' response only to the current price. The results of these studies, such as the findings from the RAND Health Insurance Experiment (HIE), are likely to be biased because they assume a "spot price" of care when calculating demand elasticities (see Manning et al., 1987; Manning and Marquis, 1996).

Some earlier theoretical and empirical work, such as Keeler et al. (1977), Ellis (1986), and Ellis and McGuire (1986), has considered consumer behavior as a sequence

---

changes since the copayment amount implies a set dollar price while coinsurance only sets the percentage of expenditures paid by the consumer, making the amount actually paid dependent on the expenditures.

of decisions about purchasing medical care in the presence of multi-tiered prices. These studies emphasized that the demand for medical care cannot be viewed as a static decision with a linear budget constraint, but must be modeled as a dynamic decision with a tiered pricing schedule for medical care. The empirical evidence of this behavior has been somewhat scarce. Recent studies by Aron-Dine et al. (2012) and Kowalski (2009) that investigate the price responsiveness of medical care using medical claims data provide evidence that consumers indeed may take into account the variable price of care when their health insurance policies include a deductible.

The objective of this research is to empirically examine the demand for medical care when a consumer faces a multi-tiered pricing structure such as a health insurance policy with a deductible. By building on the existing theoretical and empirical studies, this paper will explore whether consumers respond to the nonlinear structure of their health insurance contracts by examining which price of care, current or expected end-of-year, they use when choosing how much medical care to consume within an accounting period (typically, one calendar year).

This topic fits well into a broader area of research in economics—the aspects of consumer behavior manifested as either myopic (short-sighted) or forward-looking. The distinction between the two types of behavior is in the understanding of which price, current or expected future price, an individual uses to determine her current demand for goods, given both anticipated and unanticipated future expenditures. Studies in various areas have examined the presence of this type of anticipatory effects in consumer behavior. For example, Chevalier and Goolsbee (2009) investigate whether consumers of durable goods are forward-looking by looking at the demand for college textbooks. Using the nonlinear structure of insurance contracts as a proxy for the change in prices during one accounting period may serve as a good empirical platform to investigate whether consumers are myopic or forward-looking in their demand for medical care.

This study contributes to the existing literature and discussions in several ways. First, it examines how consumers respond to the nonlinear pricing structures of their insurance contracts. Secondly, it investigates the anticipatory effects in the demand for medical care. It accomplishes both tasks via an innovative empirical setting by focusing only on one subsample of the insured population: pregnant women who face a deductible in their insurance policy. Some of these women give birth during different months of the same calendar year, while others give birth in the next calendar year. Eliminating those with complicated pregnancies, the remaining subsample provides a quasi-experimental setting for testing the hypothesis of myopic vs. forward-looking behavior in the demand for medical care. It is expected that all pregnant women, regardless of the first month of pregnancy, will accrue certain medical expenditures, most of which will be predictable, thus reducing the need to account for the uncertainty in expenditures in both the theoretical model and empirical strategy. However, the timing of labor matters for the expected end-of-year price of care: those who are expected to deliver before the end of the calendar year (and hence incur the costs of labor and delivery, which exceed the deductible) are likely to face a lower future price because of the certainty of meeting the deductible versus those who do not give birth in the same year and may or may not reach their deductible in that calendar year.

This paper presents two ways to identify forward-looking behavior. The first, described in Chapter 6, uses the fact that the costs of labor and delivery exceed the deductible with certainty and thus lower the expected end-of-year price for the women in the labor group. This is an innovative approach that is based on a large exogenous health event that moves a consumer onto a different part of the nonlinear budget constraint. For this approach, all annual non-pregnancy related expenditures and utilization for pregnant women are examined to see whether and how giving birth affects their demand for these services and which prices, current or expected, they use. The main assumption here is that labor changes the price for all non-pregnancy related care, so women who give birth are expected to purchase more of this care throughout the calendar year. Several empirical approaches are used to examine this demand, applied to both expenditures and utilization. The results show that labor lowers the expected future price and increases the demand for non-pregnancy related care; however, other factors related to health, such as age, have a stronger effect on demand for medical care for the labor group than the price-lowering impact of labor alone. Mixed evidence is found in regards to forward-looking behavior: while there is some evidence that women indeed use their anticipated future prices when making purchasing decisions, full myopia is also rejected.

Chapter 7 applies a more established approach to examining demand under nonlinearities and follows the methodology used by a few recent studies (Jung et al., 2014a and b; Aron-Dine et al., 2012): it constructs and incorporates the expected-end-of-year price explicitly into the demand equation by estimating the probability of hitting the deductible for each woman. The demand for care is then examined using the estimated price. Annual demand, expenditures and utilization, is examined as the outcomes. The robust findings indicate that those who reach their deductible and hence face a lower expected end-of-year price have higher demand for care. The significant response to the expected future price provides further evidence in favor of forward-looking behavior and also highlights the importance of considering the nonlinearities in price when investigating the demand for medical care.

Chapter 8 uses both approaches, labor as a proxy for the expected price and using the probability of reaching the deductible to estimate the expected price, but applies them to a subset of the full estimation sample and examines the response to price during shorter spending periods. For a smaller subsample, it is possible to identify expenditures during different trimesters, as well as total non-pregnancy related spending that happened during pregnancy. Isolating the discretionary spending that happened during the pregnancy period, and during shorter intervals during pregnancy (i.e. during separate trimesters), levels the "playing field" for women in the labor and non-labor groups and eliminates both the potential unobserved health heterogeneity and the possibility of capturing an effect other than price when considering all annual expenditures, regardless if they happened during pregnancy or not. The findings in this Section corroborate the findings in both Sections 6 and 7. Women use the EFP when determining their spending during pregnancy, and this impact is more pronounced under the second empirical approach that includes EFP in the demand equation. Giving birth in a calendar year has a positive effect on spending through lowering the price, but a negative effect as a health covariate.

This work provides a different approach to the empirical research on nonlinear pricing undertaken in the existing literature. First, by looking at a large anticipated health event, such as labor and delivery, the uncertainty of meeting the deductible is eliminated. The price change is predictable and anticipated, in a fairly homogenous health setting. Additionally, this scenario requires fewer assumptions about the nature of health spending and utilization by eliminating non-emergency care. The uncertainty of exceeding the deductible and the isolation of more discretionary medical care have not been dealt with sufficiently in the previous work. The examination of demand for more discretionary care, which can be isolated in the study sample, is a unique grounds for testing for anticipatory effects in the context of nonlinear pricing. Price elasticities calculated in this scenario also provide more insight into the nature of demand for medical care.

This study is also a timely contribution to the research and policy discussions because it is one of the few to use a correct concept of price to look at the demand for medical care in the presence of cost-sharing thresholds in health insurance contracts. Since price elasticities of services are taken into account in the design of health insurance policies, it is very important to get the prices "right" and calculate elasticities based on the true prices that consumers face when making healthcare consumption decisions. Moreover, this research highlights the importance of considering the change in consumer behavior not just in response to the actual price change, but in anticipation of the changes. Labor and delivery is not only a large health event that changes the price of medical care, but an expected one, so it is reasonable to assume a behavioral response to both the price change and the anticipation of it, both of which would affect the demand for care. Similar behavior can be observed in regard to the Medicare Part D "doughnut hole" (see Jung et al., 2014 a and b).

This study has potentially important implications for policy and practice. While not setting out to find the optimal deductible, it will contribute to the research and discussions about the rationale for including cost-sharing arrangements in insurance contracts by empirically testing whether consumers respond to the nonlinear pricing schedules created by deductibles in a way anticipated by the designers of insurance policies. Additionally, knowing how consumers behave as a result of the structure of contracts and respond to expected future prices may help to answer questions about the welfare implications of health insurance by examining how valuable consumers perceive it to be in determining their consumption of healthcare.

The following chapter discusses the relevant literature on health insurance that underscores the significance of the proposed research. The theoretical model is laid out in Chapter 3. Chapter 4 discusses the sources of data and the sample selection procedure. The fifth chapter presents the overview of empirical approaches and describes three different approaches to testing the hypotheses. Chapters 6-8 are the three empirical chapters and discuss the three approaches in detail followed by the presentation and discussion of the results. Chapter 9 provides conclusions and implications for policy and research.

Chapter 2

Literature Review

The foundations of the current design of health insurance contracts in the United States are rooted in several important works by well-known economists. Traditionally, most of the research in this area has focused on balancing the tradeoff between reducing the financial risk from incurring unexpected medical expenditures and the welfare loss from inappropriate incentives to purchase medical care at an artificially low price created by the presence of insurance. The first milestone paper on health insurance focused on its role in reducing financial risk. In his 1963 paper Arrow argued that because of a high degree of uncertainty in the demand for medical care, insurance against medical expenses leads to large social gains, even though there may be some welfare loss from moral hazard. He proposed several ways of dealing with moral hazard, mostly by creating appropriate incentives for the providers (physicians) and payers (insurance companies) while offering full-coverage insurance to the consumers, with the possibility of some minimal co-insurance rate.

The second seminal piece on health insurance marks the beginning of the discussions about the optimal structure of insurance contracts to create the correct incentives for the consumers, and not the providers or payers. In response to Arrow's paper, Pauly (1968) focused on the moral hazard side of the health insurance story and argued that the risk-reducing role of insurance (Arrow's argument) was overstated and valid only when the demand for medical care was inelastic. However, any presence of elasticity in the demand for medical care (i.e. the typical downward sloping demand curve) would lead to welfare loss in the presence of the indemnity-style insurance contract, which entailed a zero price for the consumer. Thus, Pauly was the first to argue for the need to impose certain cost-sharing arrangements on the consumer to reduce welfare-decreasing moral hazard. The traditional cost-sharing methods include deductibles or coverage ceilings, co-insurance rates, and stop-loss thresholds (Feldstein, 1973); they became a permanent feature of insurance contracts in the US in the 1970s.

Since Pauly's argument, subsequent theoretical and empirical studies have adopted the notion that cost-sharing arrangements may be socially optimal and proposed various co-insurance rates that would correct the price distortion caused by the presence of insurance while taking into account the tradeoff between moral hazard and risk bearing. Feldstein (1973) proposed a co-insurance rate of up to 0.67, while others showed that an optimal insurance policy would consist of a mix of various cost-sharing arrangements (Zeckhauser, 1970) or depend on the marginal benefit for a specific medical condition (Pauly and Blavin, 2008).

The presence of cost-sharing elements in insurance contracts became the basis for the famous RAND Health Insurance Experiment (HIE) conducted in the 1970s, the results of which laid the foundation for the current healthcare policy and insurance design in the US. The participants were randomized into five groups by the degree of cost-sharing arrangements (ranging from full indemnity-like policy to a 95% co-insurance rate with a maximum out-of-pocket amount of $1000).[1] Using the cost-sharing rate as the out-of-pocket "price" of insurance, the price elasticity of demand for medical care was estimated to be -0.2 (Manning et al., 1987). Based on the results of experiment, Manning and Marquis (1996) estimated that a 0.45 coinsurance rate with no stop-loss is optimal.[2]

However, most of these theoretical and empirical works relied on a static model of consumer behavior—utility maximization in a single-period framework—even though in the presence of cost-sharing, the decision-making happens in a dynamic setting (Keeler et al., 1977, Ellis, 1986). Moreover, these studies also considered a linear budget constraint by including only the current, price for medical care, even though using a constant price

[1] Note that the last policy in the RAND HIE is similar to the current high-deductible insurance plans.

[2] Aron-Dine et al. (2012) show that the price elasticity obtained from the RAND HIE studies was in fact derived from a combination of experimental data, econometric modeling and additional statistical assumptions about out-of-sample predictions, and thus may not be accurate or robust.

(either average or marginal) of care leads to biased estimates. For example, Manning et al. (1987) based their estimations on the single price of medical care—the co-insurance rate of the group that the participants were assigned to in the RAND HIE. They refer to it as the "first-dollar approach." However, cost-sharing implies that the budget set is nonlinear as consumers face a different price for medical care depending on the structure of their insurance contract, and the first-dollar price is not constant throughout the year. In reality, a consumer with a policy that includes a deductible, a co-insurance rate, and a stop-loss (or maximum out-of-pocket limit) faces three different prices at different times during the enrollment period: full price of care before the deductible is reached, the coinsurance rate after the deductible is met, and, after the stop-loss is met, a zero price. The dynamic nature of this type of pricing schedule comes into play because the decision to purchase medical care depends on where the consumer is on the nonlinear budget constraint, and this point depends on his past consumption of care; since medical expenditures are often not predictable, decision-making is further complicated by the uncertainty of future consumption.

Indeed, most insurance plans offered by employers in the US have cost-sharing elements. According to the Kaiser Family Foundation "Employer Health Benefits" report (2012), about 77% of PPO plans, which is the most popular plan type, have an annual deductible for single coverage, with the average deductible at $733. The number of plans that feature a deductible is rising. In 2006, only 10% of firms offered plans with a single coverage deductible of $1,000 or more, compared to 34% in 2012. Most covered employees also face additional cost sharing (either coinsurance or copayment) for inpatient care (Kaiser Family Foundation, 2012). Thus, considering only a single price of care at any given point on the budget constraint, will lead to biased estimates of demand elasticities of medical care.

Even though the major studies of health insurance did not account for these nonlinearities and the dynamic nature of the demand for medical care, a few studies from the 1970s and 1980s looked at how consumers may respond to the existence of various prices (current and future) for medical care. Keeler et al. (1977) developed a theoretical model of a consumer who faces a price that varies with the number of units of care bought (such as with an insurance policy with a deductible) as well as uncertainty about the medical expenses left for the rest of the accounting period (usually 1 year). Their argument is that since expenses toward a deductible accumulate over time, even if the expenditure on today's illness does not meet the deductible, it may change the price of a future illness, which affects today's demand for medical care. Since future illnesses are unpredictable, the decision problem facing the consumer is not a single-period problem, so the decision to purchase medical care is a sequential decision. Their main contribution is the notion of an "effective price," which they define as the shadow price of one more unit of medical care given the presence of a deductible. Their model has important implications for empirical studies by raising questions of which price (marginal, average, or other) should be used to estimate the demand for medical care.

Using a dynamic model to account for sequential purchases of medical care, Ellis (1986) showed that the expected end-of-the year price for care is a good approximation to the "effective price" proposed by Keeler et al. (1977), both in situations with certainty and uncertainty. His main idea is that current expenditures have a pecuniary effect on future insurance coverage since spending an extra dollar on medical care in the current period is equivalent to reducing the other thresholds (deductible and stop-loss) by one dollar for the rest of the year. The amount of this reduction depends on the time remaining in the year and the distribution of health risks a consumer faces. He tests this model through a simulation exercise and finds that the expected end-of-year price is a valid proxy for the true shadow price of care in the presence nonlinear of price schedules as it closely tracks the shadow price under different insurance and health shocks scenarios.

Nonlinear budget sets have been considered by researchers in other fields of economics. Some of the earlier theoretical and empirical work was done on the nonconstant prices of electricity, where the marginal price of additional KWH's of electricity decreases at a given number of points determined by the total demand (see Houtthakker, 1951; Hausman et al., 1979); these studies have been enhanced by the recent empirical work on residential electricity use (Borenstein, 2009). The demand for durable goods (see Hausman (1979) for air conditioners and Chevalier and Goolsbee (2009) for textbooks) and two-part tariffs (such as telephone service—see Hausman and Trimble, 1984) are also analyzed in the context of a nonlinear budget set. The majority of the work in this area comes from labor economics that examines labor supply under progressive taxation (see Burtless and Hausman, 1978, Blundell and MaCurdy, 1999, and Saez, 2010, among others). An important methodological contribution to this literature comes from Hausman (1985) where he develops methods to deal with difficulties in modeling labor supply under nonlinear budget sets. He also develops an econometric model that explicitly incorporates uncertainty, using the case of disability insurance as an example.

A few recent empirical studies apply a similar nonlinear dynamic framework to test for the presence of anticipatory effects in the consumption of medical care. In addition to conceptualizing the nonlinear budget set theoretically, these studies propose a range of empirical methods to include the entire nonlinear budget constraint in estimating the demand for medical care. These studies use insurance claims data for employer-sponsored plans or enrollment and claims data from large employers.

An earlier study by Keeler and Rolph (1988) used the data from the RAND Health Insurance Experiment (HIE) and built on the theoretical framework proposed in Keeler et al. (1977). Based on the structure of the insurance plans in the HIE, they used the coinsurance rate and the maximum dollar expenditure (MDE) threshold to analyze whether individuals respond only to the nominal out-of-pocket price (i.e. coinsurance). They do not consider the uncertainty of illness to be a significant factor in this type of decision-making since an economically rational consumer is expected to take advantage of the varying prices and spend more when the price is low. They found that people in all plans in the HIE responded myopically as their price of care changed throughout the year, in all plans in the experiment.

Aron-Dine et al. (2012) use medical claims data from several large employers to investigate whether and to what extent individuals respond to the expected future price of care. The authors argue that the estimates of price elasticities of medical care using only the spot (current out-of-pocket) price of care, such as Manning et al. (1989), are misleading because a plan that includes cost-sharing arrangements has a nonlinear pricing structure, not just a single spot price. The premise for their empirical analysis is that a fully myopic consumer (i.e. an individual who bases consumption decisions only on current prices) would respond to the introduction of a deductible as if the "price for the first unit" of care sharply increased by a 100%, whereas a forward-looking individual who anticipates that health expenditures would exceed the deductible over the course of the year would not change behavior in response to the presence of a deductible. They use Ellis's expected end-of-year price as one of the measures of the anticipated price change, and the month of joining the insurance policy as another, and reject the null hypothesis of the fully myopic behavior in healthcare consumption. Their main empirical strategy is through sample design: they select individuals who join the same insurance plan at different times of the year.

Kowalski (2009) uses a model of a nonlinear budget setting to estimate the price elasticity of medical expenditures that accounts for the differences in medical care spending among consumers that leads to the skewed nature of the expenditure distribution. In another paper, Kowalski (2013) shows how the nonlinear pricing structure of health insurance contracts can be applied to the age-old moral hazard-adverse selection tradeoff question by estimating social welfare gains and losses. Both papers use the same identification strategy: a sample of families with a family-level deductible and a family member who had an injury, and thus incurred large medical expenditures that surpassed the individual deductible threshold. Since the amount of the individual deductible contributes to the amount family deductible, in families of four or more a large health event such as an injury would result in higher likelihood of meeting the family deductible and lower the price of care for the rest of the family. She finds that the price elasticity ranges from -2.5 to -2.0 at the 0.80 quantile, which is larger than previous estimates, especially in the upper quantiles of the expenditure distribution.

Marsh (2013) also considers the nonlinearities in the health insurance contracts and presents an estimation strategy that allows her to estimate price elasticities of medical care along different points of the demand curve. The main assumption behind her approach is monotocity: healthcare expenditures strictly increase as health status worsens. This method allows her to deal with both the selection and simultaneity issues. The first issue arises because underlying unobserved health status determines both expenditure and price: sicker people use more healthcare, moving them past the cost-sharing thresholds (i.e. kinks in the budget constraint). The second problem comes from the fact that medical expenditures and price of care are simultaneously determined. Her empirical strategy hinges on choosing individuals with expenditures right around the deductible, and the final estimate is the difference between the slopes on the different "sides" of this threshold. Using a claims dataset from a large employer, she finds that price elasticities are declining from -0.26 to -0.09 along expenditure ranges. This is a different finding from previous work that aggregates elasticities over the entire spending distribution, which is often skewed.

Two recent studies by Jung et al. (2014 a and b) consider the demand for prescription drugs, and statins in particular, in Medicare Part D setting that is nonlinear in its structure because of the coverage gap. Using Ellis (1986), they conceptualize the effective price of prescription drugs under Part D policies as expected end-of-year price and use two measures to estimate it. Their first approach is to estimate the probability that an enrollee would hit the coverage gap, which is similar to a policy that has a deductible (Aron-Dine et al., 2012) or coverage ceiling (Ellis and McGuire, 1986). Their second measure is based on the differences in pre-gap and in-gap copayments, weighted by the probability of reaching the gap. For the general case of the demand for prescription drugs by Medicare Part D enrollees, they find that pre-gap spending is sensitive to the end-of-year price, supporting Aron-Dine et al.'s (2012) finding of forward-looking behavior. In the case of statin consumption, they find that the presence of the coverage gap decreases statin compliance prior to the gap, once again finding evidence that individuals respond to the expected end-of-year price (Jung et al., 2014b).

This research builds on the theoretical models and the methodological strategies from these studies, but deals with the uncertainty of medical expenditures throughout the calendar year in an innovative way by looking at a subsample of the general population that has predictable medical expenditures. While Aron-Dine et al. (2012) estimate the price elasticity of demand for medical care using the expected end-of-year price, they do not empirically account for the unpredictability of medical expenditures of the individuals in their sample. On the other hand, Marsh's (2013) study considers only patients whose expenditures put them near the deductible, excluding those with large health shocks. Thus, the response to the nonlinear price of medical care may not reflect the demand elasticity under a more significant health event, so the estimates may capture only the more elastic part of the demand curve. Another potential limitation of this study design may arise from the ambiguity in the different responsiveness by those whose spending exceeds the deductible and those who did not, potentially biasing the results. The method used in her study does not account for unobserved heterogeneity that is unrelated to health status that may have caused some individuals to exceed the deductible while the expenditures of others remained below the threshold.

This study aims to overcome these limitations by estimating demand when much of this uncertainty and some of the unobserved heterogeneity is eliminated by the choice of the study population—pregnant women, i.e. fairly homogenous (health-wise) individuals with a similar large health event. In addition, the health shock considered here, labor and delivery, is predictable, which creates a unique context to look at price responsiveness and forward-looking behavior. Finally, the study design and construction of the estimation sample eliminate the need to explicitly estimate the expected end-of-year price. Jung et al. (2014 a and b) and Aron-Dine et al. (2012) estimate this price by calculating the probability of reaching the kink in the nonlinear budget constraint—either hitting the Part D coverage gap or reaching the in-network deductible in an insurance plan. The authors of these studies admit that the explicit estimation of the expected price is "noisy." Since women who give birth are certain to exceed the deductible, this empirical setting allows to bypass the step of estimating the end-of-year price by incorporating it implicitly. However, in Chapter 7 the expected end-of-year price is calculated explicitly using the methods from Aron-Dine et al. and Jung et al., to compare with the results from Chapter 6.

Chapter 3

Conceptual Framework

As described above, a typical private insurance contract in the US includes several components such as a deductible, a coinsurance rate, and a stop-loss, that create a nonlinear price schedule based on reaching several critical thresholds in one accounting period (usually one calendar year). The first threshold is the amount of the deductible that is determined by the policy chosen by a consumer. After the deductible has been met, the co-insurance rate is applied to the subsequent medical expenditures until the second threshold, the stop-loss, or out-pocket maximum, has been reached. Expenditures exceeding the stop-loss are fully covered by the insurance company. This 3-tiered price structure can be thought of as:

$$p(z_t) = \begin{cases} 1 & \text{when} \quad 0 < z_t \leq d \\ c & \text{when} \quad d < z_t \leq s \\ 0 & \text{when} \quad s < z_t \end{cases}$$

where z_t is accumulated expenditures on medical care, d is the deductible, c is the co-insurance rate, and s is the stop-loss. For example, if a consumer has a policy with a \$500 deductible, a 20% coinsurance rate, and a \$5,000 stop-loss, she would pay full price for care until her expenditures reach \$500 and 20% for all expenditures afterwards while the insurance policy pays the remaining 80%; once her out-of-pocket spending reaches \$5,000, she pays 0 for all other medical care consumed in the same calendar year. Note that using this example, the total amount of care consumed (paid for by the patient and the insurance) before reaching the stop-loss is \$22,500.

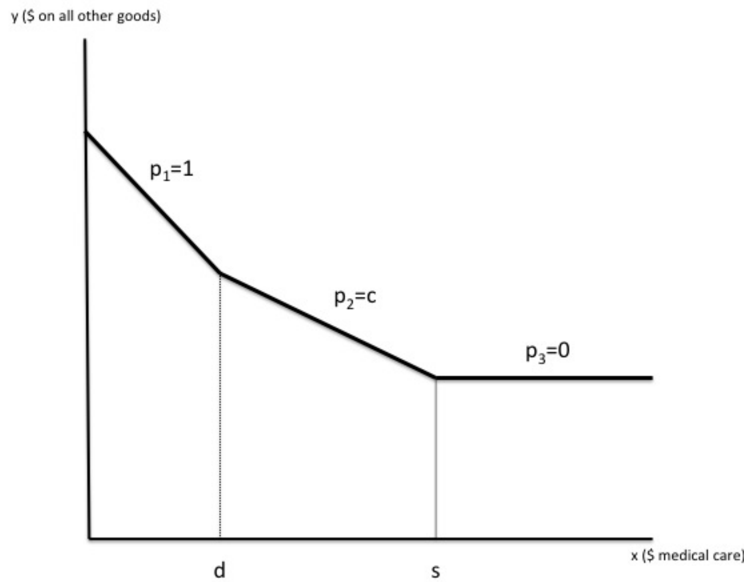

Figure 3.1: Nonlinear Budget Set

This example shows that a consumer no longer faces a typical convex budget set with a single price for each good. In the 3-tiered pricing structure, an individual has several kinks in the budget set that correspond to the critical thresholds described above. Figure 1 illustrates the effect of nonlinear pricing structure has on the budget set. It is clear that the price that the individual faces today (or, in other words, which segment of the budget line she is at) depends on the accumulated expenditures since the beginning of the accounting period. Hence, the consumption of medical care cannot be viewed in a one-period framework, but represents a sequential decision (Keeler et al., 1977). The expenditures below the deductible, while coming at full price, nonetheless have the benefit of potentially reducing the expected future price of medical care; in other words, spending an extra dollar on medical care today can reduce all critical thresholds for the rest of the year.

Another important aspect of this sequential decision-making is the presence of uncertainty about future medical expenditures. At time t, a consumer does not know what future urgent and non-urgent medical care she may demand. Thus, the appropriate formal model for this type of decision-making is a finite-horizon stochastic dynamic programming problem. Since the deductibles and stop-losses are reset at the end of the accounting period, a year is a natural horizon for making healthcare expenditure decisions in the presence of nonlinear price structures. First, a number of simplifying assumptions are made. Following Keeler et al. (1977) and Ellis (1986), I assume a risk-neutral consumer. Secondly, I assume a 2-tier pricing structure: p_1 and p_2 from Figure 1, i.e. medical expenditures and out of pocket payments that are smaller than the stop-loss amount.

In each period t consumer allocates some of her total wealth W_t to the purchase of two goods: medical care, x_t and other goods and services y_t. The price of y_t is normalized to 1. Following the well-known Grossman model of health capital (Grossman, 1972), consumer does not derive utility directly from medical care, but rather from the stock of health at the beginning of t:

$$U_t = U_t(H_t, y_t)$$

The stock of health H_t is a function of health status and the medical expenditures in the previous period as well as a health shock in the previous period, so the production function for the health stock is:

$$H_t = h(H_{t-1}, x_{t-1}, \epsilon_{t-1})$$

Using macrons over variables to denote the end-of-year values, the end-of-the year (i.e. December 31st) stock of utility is:

$$V_{T+1} = V_{T+1}(H_{T+1}, W_{T+1}, z_{T+1}) = V_{T+1}(H_{T+1}, W_{T+1}, 0^1) = \bar{V}(\bar{H}, \bar{W})$$

Thus, following Ellis (1986), the consumer's problem is

$$\max E\left[\sum_{t=1}^{12} \beta^t U_t(H_t, y_t)\right]$$

[1] Expenditures in the next enrollment period do not count towards reaching the cost-sharing thresholds in the current year.

subject to

$$
\begin{aligned}
H_{t+1} &= h(H_t, x_t, \epsilon_t) \\
W_{t+1} &= W_t - y_t - \int_{z_t}^{z_{t+1}} p(z)dz \\
z_{t+1} &= z_t + x_t \\
V_{T+1} &= \bar{V}(\bar{H}, \bar{W})
\end{aligned}
$$

where β is the personal discount factor. To simplify, I assume that $\beta = 1$.
The Bellman equation is

$$V_t(H_t, W_t, z_t) = \max\left[U_t(H_t, y_t) + EV_{t+1}(H_{t+1}, W_{t+1}, z_{t+1})\right]$$

Ellis (1986) discusses why the nonlinear structure of $p(z)$ creates discontinuities that complicate the analytical solution to the maximization problem. However, he proves a result about the equilibrium condition for this problem. He shows that if the expenditures on both x_t and y_t are positive, then the marginal rate of substitution between x_t and y_t should equal to the expected end of the year price:

$$\bar{p}^e = \frac{E\left[\frac{\partial V_{t+1}}{\partial H_{t+1}} \frac{\partial H_{t+1}}{\partial x_t}\right]}{\frac{\partial U_t}{\partial y_t}} = E[p(\bar{z})] \tag{3.1}$$

This price provides a good analytical equivalent to Keeler et al.'s (1977) effective price and reflects the true price of medical care in the presence of a health insurance policy with a deductible. The expected end-of-year price depends on the medical expenditures accumulated throughout the year, which are often unpredictable. Thus, the general population with a health insurance policy that includes a deductible has to infer the end-of-year price, given the expenditures anticipated throughout the year. Hence, time and stock of health, as well as health shocks after the current period, are important factors in predicting the end-of-year price, so the effective price can be thought of as

$$\bar{p}^e = f(z_t, H_t, t, \epsilon_{t+1}) \tag{3.2}$$

This equation implies that in order to find the effective price with some degree of precision, we have to account not only for the health shocks, but for their time and size to determine whether the accumulate expenditures would exceed the deductible, i.e.

$z_{T+1} > d$. However, if we could isolate a subsample within the general population with accumulated healthcare expenditures that would exceed the deductible with a degree of certainty, we could eliminate some of the uncertainty associated with predicting the end-of-year price. This would be a group with a "serious" medical condition that is likely to get expensive treatment. If such a subset of population were available, we could use the price of care after the deductible is reached (p_2 from Figure 1) as the end-of-year price.

One group that fits this qualification is a cohort of women whose delivery date falls within the same calendar year. The Agency for Healthcare Quality and Research (AHRQ) estimates show that in 2010 the average total (out-of-pocket and insurance) cost of labor and delivery ranged from $10,166 for vaginal births with no complications to $23,111 for Cesarean births with complications. On the other hand, the Kaiser Family Foundation finds that the average deductible in the US private policies is slightly below $1,000. Therefore, for any woman giving birth, the costs of labor and delivery will certainly exceed the deductible.[2] Thus, for this group of individuals we can predict their end-of-year price with a reasonable degree of certainty, using the structure of their insurance contract.

One the other hand, women who are pregnant in the same calendar year, but will deliver during the next calendar year, are uncertain about their medical expenditures for the current year; given their health status at t, time to the end of the year $T - t$, and health shocks in future periods of the same calendar year ϵ_{t+1}, their accumulated health expenditures may or may not meet the deductible. Thus, they are less likely to be able to predict their expected end-of-year price. The main hypothesis being tested is whether women who give birth within a calendar year and thus will certainly exceed their deductible will consume more medical care during that year than women who are pregnant but will not give birth until the following year. Using both groups of pregnant women in the same estimations of medical care consumption can help shed light on whether and how consumers respond to the anticipated end-of-year price.

[2] Under most plans with a deductible, any inpatient expense such as labor and delivery is subject to deductibles.

Chapter 4

Data and Sample Selection

4.1 Data Description and Sample Construction

The data for this study come from the MarketScan Commercial Claims and Benefit Plan Design Databases compiled by Truven Health Analytics and accessed through the National Bureau of Economic Research (NBER). These databases capture person-specific clinical utilization and expenditures on inpatient, outpatient, and other medical services, as well as insurance plan enrollment from approximately 45 large employers, health plans, and government and public organizations. The MarketScan Databases link paid claims and encounter data to detailed patient information across sites and types of providers, and over time. The Commercial Claims database has information on active employers and their dependents insured by employer-sponsored plans. The Benefit Plan Design Database contains data abstracted from the summary plan descriptions of selected insurance plans represented in the Commercial Claims Database (MarketScan Database, Thompson Medstat, 2004).

These data are well-suited for the analyses in this study. The merged dataset allows the identification of pregnant women, the date of their labor and delivery, and any complications, as well as their utilization of inpatient and outpatient services and payments made by the enrollee and insurer. Moreover, according to the 2012 Kaiser Annual Survery of Employer Health Benefits, 81% of employed persons in the U.S. enroll in health plans offered by their employer, so the results obtained using these data are fairly generalizable. The main limitation of the dataset is that only a few demographic and employment-related characteristics such as age, employment status, and industry of employment are present. However, since the entire sample comes from a pool of insured employed individuals, having socioeconomic variables may not be critical for this study because some of this observable heterogeneity is eliminated.

Finally, this data set is especially appropriate for a study using this particular subsample of the population. Most states have generous Medicaid eligibility thresholds for expectant mothers, ranging from 133% to 300% of the federal poverty line. Therefore, women in the lower income bracket may choose to forgo their employer-sponsored insurance plan in favor of Medicaid coverage while they are pregnant. Hence, the pregnant women who remain on their or their spouse's plan are expected to be in the income range that is above the Medicaid eligibility level, and hence homogenous to some degree in terms of socioeconomic characteristics. In addition to controlling for income heterogeneity with the sample of women covered by an employer-sponsored health plan, the absence of the income variables from the data set becomes even less problematic because of the nature of the health condition considered in the study.

The data from both databases were matched for 1996-2009. The sample used in the analyses was carefully selected by the following steps:

1. First, the sample was limited to individuals enrolled in non-managed (non-HMO) care plans because the managed care plans typically do not have cost-sharing elements (Kaiser Foundation, 2012).

2. Women of childbearing age (18-44) were selected from the Commercial Claims Database; the available data include inpatient and outpatient medical care use and expenditures, some demographic characteristics, and plan enrollment information. In addition, the sample of women was limited to full- or part-time employees.

3. Using the ICD-9CM codes, pregnancy and pregnancy complications were identified in the services claims files.

4. With guidance from the clinical literature (such as Kuklina et al., 2008), labor was identified using diagnoses and procedure codes (both ICD-9CM and CPT codes). Both outpatient and inpatient claims were used to identify labor to account for the outpatient delivery centers (midwifery clinics and birthing centers) as well as the possibility that an inpatient delivery was initiated with an outpatient visit.

5. Inpatient and outpatient files were merged with the plan enrollment files and the Benefit Plan Design files. Subsequently, only plans that began on January 1st of each calendar year were kept in the sample.

6. The sample was then limited to pregnant women with normal pregnancies, excluding pregnancies with complications that affect the management of the mother.

7. The final sample includes normally pregnant women enrolled in a preferred provider organization (PPO) plan with a fixed non-zero deductible.

Another exclusion was made from the final dataset: women who gave labor but had total out-of-pocket (OOP) spending that was less than their individual deductible. These women constituted only about 1.8% of the final sample, and about 10% of them had zero OOP expenditures. Since labor and delivery expenses are usually subject to deductible, women who give birth should have OOP spending that is at least equal, and mostly larger, than their deductibles. Therefore, those who did not exceed their deductible may have had coverage through other sources (spouse's insurance policy, Medicaid, etc) or an atypical policy that would include coverage of pregnancy-related inpatient and outpatient care that are not subject to a deductible. For these women, the costs of labor and delivery may not result in the anticipated price reduction and thus have a different impact on the expected end-of-year price.

4.2 Selection Issues

Given the research question and the sample construction, it is not unreasonable to suspect the presence of plan switching in anticipation of large expenditures associated with giving birth, which could be a indication of adverse selection and thus bias the results. Selection may arise from different kinds of decisions about a health insurance plan. First, a woman may choose to switch into a zero deductible plan in the year of labor (in anticipation of having to pay out the deductible) if one is offered by her company. Another decision may be the type of plan chosen for the year of labor. The dataset obtained after Step 6 of the cleaning procedure (before limiting the observations to PPO plans) includes claims from comprehensive plans, exclusive provider organization (EPO) plans, point-of-service (POS) plans, preferred provider organization (PPO) plans, and consumer driven health care (CDHP) plans. About 65% of all women between 18 and 44 years of age in the data are enrolled in a PPO plan, followed by 23.5% in POS plans. The rest are distributed across the other non-managed care plans. HMO and POS with capitation plans, the two types of managed plans in the dataset, comprise less than 1%.[1]

A woman may choose to switch into a more "generous" plan type in anticipation of labor. The third source of switching may come from selecting a lower-deductible plan in the year of labor.

To investigate this further, the patterns in switching between plans were examined in the general sample of women (pregnant and non-pregnant) and among the pregnant women in the sample. Those who were in the sample for several years consecutively were kept for this exercise. Since not every woman who was pregnant in one year of the data remained in the data during the year of labor, the data on consecutive years of pregnancy and birth are available only for a subsample of pregnant women and are used in the analysis of switching behavior. Table 4.1 summarizes the results of the investigation of switching behavior. The first three columns refer to the analyses on the sample that included all women of child-bearing age. Pregnant women and women who gave birth while in the sample for two consecutive years (labeled "birth") are examined in comparison to the general sample. The last two columns reflect the analyses performed only on the women who are pregnant; some of this cohort gave birth and others did not. Since not all women in the sample may be in the plan in consecutive years, the sample used in the selection analyses was further limited to women who were in the data in consecutive years.

The results show that on average women who give birth are more likely to switch into a zero-deductible plan—around 10%, compared to 5% among all women. This could lead to biased results because it would be difficult to distinguish whether the response to higher prices happens when choosing a health plan (i.e. decision made before the start of the enrollment period) rather than to the price of care under a certain health plan (decision made during the enrollment period). To deal with this source of selection, only those in non-zero deductible plans are kept in the final sample. The exercise also

[1] Switching into a managed care plan may be another decision that a pregnant woman may make in anticipation of labor, since managed care plans typically do not include a deductible. This potential switching was examined and was only found to occur in less than 0.001% of cases. Overall, managed care plans were excluded from any type of analysis in this study.

Table 4.1: Summary of Selection Analysis

	All Women			Pregnant Women	
	All	Pregnant	Birth	All	Birth
Switched into 0-deductible plan (%)	6.5	5.96***	10.29***	5.96	10.34***
Switched between plan types (%)	43.72	41.51***	8.59***	41.51	8.84***
Increase in size of deductible (%)	88.88	143.79***	11.63***	143.79	11.37***
Proportion with decrease in size of deductible (%)	0.67	1.81***	4.45***	1.81	4.59***
Observations	1,628,545	23.12%	3.63%	376,483	14.46%

Note: Statistical testing is done for mean equality between two groups (e.g. birth=1 and birth=0).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

revealed that a high number of women switched between plan types; moreover, this trend of switching between plan types was similar in the the general sample of women and in the sample of pregnant women, so pregnancy and expected expenditures associated with it may not be the primary reason for changing insurance plans. However, this type of switching is significantly lower for women who give birth and hence may not be a concern here. To deal with potential heterogeneity in insurance coverage, only those enrolled in a PPO plan were kept in the final sample.

Finally, there are significant differences in the changes in the size of the deductible between the general sample of women and those who gave birth. The increases in the deductible are much smaller for the latter group, indicating that they may remain in the same insurance plan (and the increase in the deductible is attributable to yearly increases built in the health plan offerings). Moreover, the proportion of women for whom the deductibles got smaller is larger among those who gave birth, indicating that they indeed may switch to a plan with a lower deductible more so than other women, but that proportion is not large: about 4.5%. Furthermore, this issue may be negligible here because the study is looking at the within-period variation in the consumption of medical care, i.e. decisions made within the benefit period, and specifically on whether a pregnant woman gives birth within that benefit period and how this change in expected end-of-year price affects her behavior within that specific period. Even though the data are pooled across years, the behavior of interest occurs within each calendar year. The costs of labor and delivery would exceed any individual deductible; this fact is at the heart of the identification strategy used in the paper. The size of deductible is used as a proxy for the current price, as explained in Section 5, but should not play a role in consumption decisions if consumers are forward-looking. Switching within the benefit period is controlled by selecting only those who are enrolled in the same plan for the entire year.

The limitation of these exercises to detect switching patterns is that given the data, it is only possible to examine all types of switching—both voluntarily by the enrollee or mandated by the employer as a result of new insurance plan offerings. It is noteworthy that according to the 2012 Kaiser Annual Survey of Health Benefits, the number of enrollees grandfathered into their original plan has been declining—only about 48% of employees in 2012 remained in the same type of insurance plan as in the previous year. Thus, the yearly differences in plans in the sample may be part of the larger structural trend where enrollees switch plans involuntarily as a result of the changing insurance offerings by their employer.

Additionally, unrelated to plan selection and switching, bias may arise from observed and unobserved health heterogeneity. One source of the health-related bias comes from the health status unrelated to pregnancy. However, these trends in health would be similar among all pregnant women, both in the labor and non-labor groups. If any of these unrelated health conditions affect pregnancy, they are typically coded as "pregnancy complications" in the claims data. The second cause of concern for the health related bias is the following: it is possible that a pregnant woman who has health issues affecting pregnancy (but unrelated to labor) may purchase more medical care regardless of the size of the deductible (i.e. whether she gives birth within that benefit period). This potential source of bias is dealt with by omitting those with any type of complications during their pregnancy. In fact, the number of women with at least one code for pregnancy-related complication is quite large (about 57%), so omitting them leads to a fairly "conservative" sample. Thus, both selection issues related to plan choice and concerns about health heterogeneity are limited through the sample construction procedures.

4.3 Descriptive Statistics

As mentioned above, the main limitation of a claims dataset is that only few demographic variables are available. However, enough variables are available to control for potential confounding effects of labor on the demand for medical care. Age and its square are expected to affect the demand for care by pregnant women since maternal age is recognized in the literature as a determinant of the overall wellbeing during pregnancy as well as pregnancy outcomes (see, for example, Seoud et al., 2002). Moreover, the interaction term of labor and age controls for potential differential impact of age (through the associated health effects) on the demand for medical care for the labor group. Family size, another covariate in the demand equation, is defined in the data as the number of enrollees in the health insurance plan and hence may not be a precise count of family members. However, it must be included to control for the effect that a larger family may have on demand for care through time and other opportunity costs. An interaction term of labor and family size is also added to account for the possibility that women who have given birth in the past may understand the price-lowering effect of labor and respond to the EFP differently than those who are in their first pregnancy. The "number of months before labor" variable is defined as "month of labor-1" and is coded to 0 for women who do not give birth within a calendar year. This variable captures the number of months that a woman has opportunities for forward-looking behavior as she anticipates the price change caused by the labor and delivery costs. It is top-coded at 9 because women who give birth later in the year were not pregnant or aware of pregnancy in the beginning of the year, and hence could possibly have at most 9 months for forward-looking behavior. Finally, a variable called "type of employee", while not shown here, is added to the model. It is a categorical variable that denotes whether an employee is a salaried, hourly, union, non-union, or a combination of these, or unknown. The distribution of this variable is not descriptive of any particular trend; however, it is a useful covariate because it can control for differences in workplace rules, maternity benefits, and even some unobserved socioeconomic heterogeneity.

The demand variables are constructed as follows. Total spending variable comprises payments to the provider or facility from all sources: insurance company payments and all out-of-pocket payments from the beneficiary (deductibles, coinsurance and copayments). Expenditures for non-pregnancy related care are constructed by subtracting all outpatient and inpatient pregnancy and labor related expenditures from the total expenditure variable. The latter are identified using available clinical codes, including the grouping codes created by MarketScan, all services and procedures that are related to pregnancy and labor (including prenatal and postpartum care, labor and delivery, antenatal screening). Spending is first aggregated by the date of service, and then added up to create annual figures. The utilization variables are constructed by identifying the type of visit by service or admission date, using the same clinical codes as for the expenditure variables. A visit is considered as non-pregnancy related if there are no codes for any type of pregnancy or labor-related care received on that visit. For example, if there are codes for prenatal care (along with other codes) posted for a certain date of service, the whole visit is identified as pregnancy related and is not counted. The visits are then summed by year to construct the annual number of non-pregnancy related visits.

Table 4.2 presents the summary statistics for the study sample, in general and by labor. About 46% of women in the sample gave birth within a calendar year, with the average month before labor at around 5, which points to a fairly uniform distribution of labor months. About half of the women in the sample are enrolled in a plan as the principal plan holder, with this proportion being slightly lower for the labor group. Age and family size are similar across labor and non-labor women. As expected, total annual spending and total utilization are higher for those who give birth due to the costs of labor. However, the labor group on average spends less on non-pregnancy related expenditures than those who do not give birth in the same year. They also have fewer visits unrelated to pregnancy than the labor group.[2]

Table 4.3 shows the characteristics of the health insurance plans represented in the sample. The PPO plans used in this study have the three-tiered structure discussed above. First, consumers pay full price for care until the deductible is reached, with the exception of preventative care. Then, a single coinsurance rate is applied to all medical care, until the out-of-pocket maximum is reached (MarketScan Database, 2004). The size of the average individual deductible is around $360, with a 13% coinsurance rate,

[2] All financial variables, such as expenditures and cost-sharing thresholds, are in 2009 dollars, having been adjusted by the medical portion of the Consumer Price Index (CPI).

and \$1800 individual OOP maximum. These numbers are fairly similar across the labor and non-labor groups, which highlights that there is no significant evidence of systematic selection into lower deductible plans by women who give birth. It is noteworthy that overall the deductibles are not very high. Since the study sample spans 14 years (from 1996-2010), these figures may reflect the evolution in the structure of health insurance policies, with higher deductibles being a fairly recent feature.

Two additional notes should be made regarding the sample used for the statistics presented here. Most of the methods used in the study are most suitable for nonnegative-valued data (see Section 5.3 for more discussion). With the exception of the subsamples of the data with high zero mass, the general sample contains fewer than 1% of women with zero expenditures, so omitting them does not pose a problem for the generalizability of results. Additionally, observations with total non-pregnancy related expenditures over $\$100,000$ are also omitted; these constitute fewer than 1% of the sample. These are omitted to further control for potential unobserved heterogeneity due to health or selection. Reasonably, those with very high non-pregnancy expenditures have worse health status, so their spending may be driven by reasons other than price response.

Table 4.2: Demographics and Healthcare Use

	All Sample	No-labor	Labor
Labor	0.46	0.00	1.00
	(0.50)	(0.00)	(0.00)
Months before labor	-	0.00	5.08
	-	(0.00)	(3.02)
Age	29.41	29.27	29.58
	(5.46)	(5.73)	(5.13)
Family size	4.20	4.15	4.25
	(2.56)	(2.57)	(2.55)
Relationship to employee	0.52	0.55	0.48
(1=self; 0=dependent)	(0.50)	(0.50)	(0.50)
Total medical spending ($)	5037.04	2026.80	8537.13
	(5873.15)	(3834.63)	(5892.19)
Total non-pregnancy spending ($)	1190.93	1404.15	943.02
	(3180.04)	(3645.63)	(2510.75)
Total medical utilization (# of visits)	10.26	9.57	11.07
	(8.15)	(8.76)	(7.29)
Total non-pregnancy utilization (# of visits)	6.97	7.23	6.66
	(7.47)	(8.24)	(6.46)
Observations	38808	20864	17944

Note: Standard deviations in parentheses.

Table 4.3: Characteristics of Insurance Plans

	All Sample	No-labor	Labor
Individual deductible ($)	363.18	372.67	352.15
	(239.02)	(271.87)	(193.37)
Family deductible ($)	823.63	840.53	803.99
	(535.75)	(590.96)	(462.60)
Copayment for primary care ($)	5.47	5.43	5.52
	(8.64)	(8.67)	(8.62)
Copayment for specialist care ($)	6.85	6.78	6.92
	(11.82)	(11.76)	(11.88)
Copayment for ER ($)	38.09	37.08	39.28
	(39.74)	(39.68)	(39.77)
Copayment for inpatient care ($)	7.07	7.61	6.44
	(41.48)	(43.26)	(39.30)
Coinsurance rate	0.13	0.13	0.13
	(0.08)	(0.08)	(0.08)
Preventative care coverage	0.85	0.85	0.85
	(0.36)	(0.36)	(0.35)
Individual OOP maximum ($)	1867.71	1887.05	1845.22
	(1097.80)	(1111.28)	(1081.52)
Family OOP maximum ($)	3593.27	3626.88	3554.20
	(2181.11)	(2194.91)	(2164.36)
Observations	38808	20864	17944

Note: Standard deviations in parentheses.

Chapter 5

Empirical Approach

5.1 Main Idea

To investigate whether individuals take into account the tiered pricing structure by using expected end-of-year prices versus current prices of medical care, a demand equation for medical care for an insured person who has a deductible in her policy is estimated. As discussed in the section above, the sample is limited to normally pregnant women who have an employer-provided PPO plan with a deductible. The reduced form of the demand equation is:[1]

$$y_{ijt} = \alpha + \beta p_i^e + \gamma \mathbf{X}_i + \delta \mathbf{Insur}_j + \tau \mathbf{Year}_t + \varepsilon_{ijt} \tag{5.1}$$

where

y_{ijt} − outcome variable

p_i^e − expected end-of-year price

$\mathbf{X}_i$ − vector of enrollee characteristics

$\mathbf{Insur}_j$ − vector of insurance plan characteristics

$\mathbf{year}_t$ − year dummy variables

ε − unobserved heterogeneity

The unit of observation is a pregnant woman i in a health plan j in year t. The outcome variable is the total non-pregnancy related expenditures or utilization within

[1] The bolded variables indicate vectors, and the unbolded stand for scalars.

a calendar year. These exclude the costs and visits associated with labor and delivery because a woman who gives birth incurs these expenditures and uses these services independent of which price, current or expected end-of-year, she uses to make decisions. As described above, prenatal and postpartum care, including antenatal screening and fertility treatments resulting in pregnancy, are also excluded because these services usually are non-discretionary and may depend on how late in the gestation period a woman is—that is, the demand for this type of medical care may not happen in response to the anticipated price change. Moreover, some health plans may consider prenatal care part of preventative care coverage, in which case it may not be subject to deductibles. It should be noted that in the demand specification given in equation (5.1), the null hypothesis is that consumers are myopic (similar to Aron-Dine et al., 2012) because the "size of the deductible" variable is included. Hence, this study allows a direct test for the presence of myopia in consumer behavior. The conclusions on the presence of forward-looking behavior can be made based on which price, current or expected end-of-year, women in the sample use when making purchasing decisions.

However, the expected end-of-year price is endogenous: it is determined simultaneously with the decision to purchase medical care. The presence of the cost-sharing thresholds, particularly the deductible, makes the price of any additional care a function of the accumulated expenditures as well as future expenditures, as shown in equation (3.2). Therefore, while the expected end-of-year price can be constructed empirically, given the data used for the study, it cannot be used directly in the demand estimations as it will lead to biased results if the simultaneity is not accounted for (Wooldridge, 2010).

This study addresses this endogeneity in two ways. The first is an innovative approach that uses a large health event such as labor that would move a woman past the deductible threshold and lower the expected future price. It is reasonable to assume that the timing of labor and delivery is exogenous to the pregnant woman: while she knows the approximate date of delivery and can anticipate the price change caused by the costs of birth, she cannot choose the year (or even the month) of labor because of the medically established norms for gestation period.[2] This is the basis for the identification

[2] It is possible that some women who are close to their delivery date may "choose" to give birth (via induction of labor or a planned c-section) before the end of the calendar year to take advantage of the

strategy for this part of the study. Labor is used as either a proxy or an instrument for EFP (based on additional assumptions) so that EFP does not appear in the estimating equation. This identification strategy lends itself to several empirical "interpretations," some of which require fewer assumptions, but all of which aim to estimate the demand equation given by (5.1).

The other approach uses a strategy previously employed in a few other studies that looked at the effect of the price nonlinearity in insurance on the demand for care (Aron-Dine et al., 2012; Jung et al., 2014a and b). This approach explicitly incorporates the expected end-of-year price in the demand equation—this price is calculated by estimating the probability of reaching the deductible. Labor is not used as the price-changing variable, but as a health covariate. Even though this approach may be a bit more "noisy" that the first approach due to including an estimated price into the demand analyses, it is useful here since it would take into account that some non-labor women may reach their deductible and face a lower future price. Note that the EFP variable, constructed as described in Figure 1 (see Section 6.3) and used in the demand equation (5.1), is not included in the estimations since it is endogenous; instead, it is empirically calculated based on the predicted probability of reaching the deductible.

5.2 Modeling Issues

Before describing the details of the empirical approaches used in Section 6 and 7 (the empirical sections), it is necessary to discuss the problems that arise when modeling healthcare data. The literature typically identifies several problems that would lead to less precise estimates of mean and marginal effects or even biased results, if not dealt with appropriately (see, for example, Jones, 2000). These issues usually arise from the nature of the data generating mechanism: they are related to the distribution. A variety of econometric techniques have been proposed to deal with these issues, although none seems to be the perfect solution or the best-in-class applicable in all situations. These commonly-identified issues, applied to the data used in this study, are discussed below. Most scholars agree that raw-scale, or untransformed, Least Squares methods do not

insurance coverage. To account for this possibility and for the possibility of a shorter period for forward-looking behavior, in the sensitivity analyses the month of labor is limited to April-September—this is presented and discussed in Section 6.6.

perform well with healthcare expenditure and utilization data because the nature of the data often violates the standard Least Squares assumptions (Manning and Mullahy, 2001; Jones, 2010; Deb et al., 2013). Three estimators proposed in the literature as best-suited for analyses involving healthcare data are log-OLS, Generalized Linear Model (GLM) with log link and gamma variance, and a newer estimator, Extended Estimating Equations (EEE), or extended GLM. A number of tests were performed to see which of these estimators may be more suited for these data and produce more precise and consistent estimates and lead to the most accurate inference.

The first issue that makes application of traditional OLS methods unsuitable is the skewed nature of the distribution of healthcare expenditures. Typically, most expenditures are concentrated in the right tail of the distribution—a few high spenders account for a large portion of total spending. One way to deal with the heavy right tail is to transform the outcome variable into logs. However, this retransformation causes problems when calculating marginal effects (Duan et al., 1983; Mullahy, 1998). The summary statistics for the continuous outcome variable, non-pregnancy related expenditures, are mean=1190.96, variance=10100000, the coefficient of skewness=11.53, and the coefficient of kurtosis=209.62. In agreement with the majority of research in health econometrics, raw-scale OLS will not be used in this study because the results will biased, inefficient, and sensitive to influential outliers (Deb et al., 2013).[3]

Figures 5.1 and 5.2 show the density function of the raw-scale expenditures and the transformed log-scale expenditures. The graphical findings confirm that the distribution is very skewed and heavy in the right tail, which calls for a log transformation in the least squared methods. The analysis of the residual from the log-OLS model show that it is close to being normally distributed: skewness=-0.132 and kurtosis=3.68. However, the Skewness/Kurtosis Test for normality fails, indicating that the Least Squares methods may not be most suitable for these data. Figure 5.3 below shows the density for the Studentized residuals with log scale. The non-normality of residual fails one of the classical OLS assumptions and may cause problems for inference and for retransformation to obtain marginal effects after log-OLS. Since log-scale kurtosis is greater than 3, GLM may also suffer precision losses (Manning and Mullahy, 2001). Additionally, health

[3] Other transformations have been used in the literature in conduction with OLS, square root being one. The Box-Cox test, which is used determine the relationship between $x\beta$ and $E(y|x)$, showed that for these data the log transformation is the closest to the true relationship.

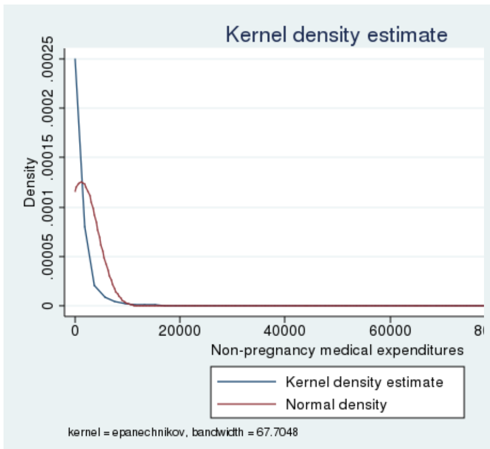

Figure 5.1: Density for Non-Pregnancy Expenditures

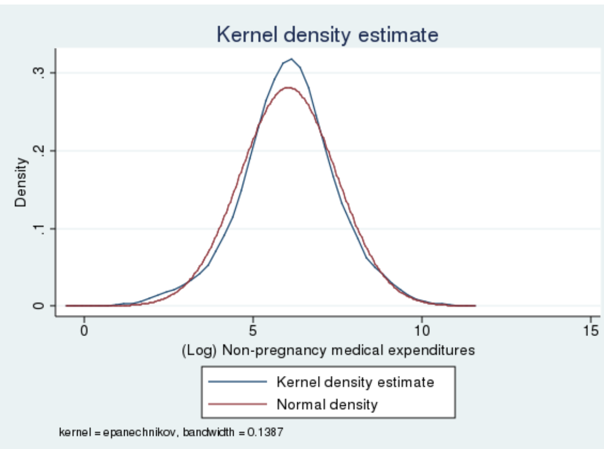

Figure 5.2: Density for Log of Non-Pregnancy Expenditures

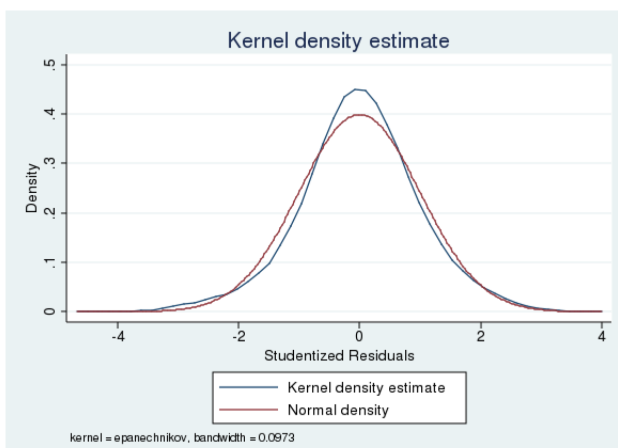

Figure 5.3: Density for OLS Studentized Residuals, log scale

expenditure and utilization data may have heteroskedasticity that can arise from several sources. The Breusch-Pagan test, a standard test for heteroskedasticity, is applied to the log-scale after log-OLS. The null hypothesis of constant variance is rejected at the 1% level. Additionally, the GLM version of the Park test, suggested by Deb et al. (2013), comes up with the same finding: these data have complex heteroskedasticity that likely arises from several sources, which makes them difficult to model. Hence, Least Squares methods may not capture the correct relationship between the mean and the variance. It makes the retransformation issues for log-OLS more complicated (see discussion below). GLM estimators would be consistent in this case, but at some loss of efficiency (Manning and Mullahy, 2001).

Another problem identified in the literature is that the response to covariates may not be linear (Deb et al., 2013). The Pregibon's Link test is applied to both log-OLS and GLM to test for linearity of response on scale of estimation. Table 5.1 shows that both estimators do not fail this test, so nonlinearity is not a problem leading to bias (Deb et al., 2013). Results of other diagnostic tests for model fit are also presented in Table 5.1. The modified Hosmer-Lemeshow test is used to determine if there is systematic bias in fit on raw scale. It is conducted on log-OLS, GLM, and the Extended Estimating Equations (EEE) model, which will be described later in this section. All three models do not fail the test, so no systematic pattern is detected. The Pearson's Chi Test is also used to assess the goodness of fit; since all three models pass it, it is assumed that the functional form of the covariates in the log-OLS and the choice of the link function for GLM and EEE are correct.[4] Finally, a modified Park test (or a GLM family) is suggested by Deb et al. (2013) to determine the relationship between raw-scale mean and variance functions. Only the gamma family is not rejected, which is the family most often used in the estimations involving healthcare data.

These tests, which are guides for the most appropriate empirical method, paint a complicated picture that confirms the conclusions in the literature: the "best" estimator does not exist for the analyses of healthcare data, and the choice of empirical method should be guided by the data. For example, the tests show that while non-linearity in covariates may not be an issue, log-OLS may create imprecise results due to complex

[4] A heteroskedastic smearing factor is applied to the log-OLS retransformed residuals to conduct these diagnostic tests—this will be discussed later in this section.

Table 5.1: Model Fit Tests

Test	Objective	Model Tested	Results
Pregibon's Link Test	To test for linearity	log-OLS and GLM	does not fail
Hosmer-Lemershow Test	Test for goodness of fit	log-OLS, GLM, EEE	does not fail
Pearson's Chi Test	Test for goodness of fit	log-OLS, GLM, EEE	does not fail
GLM Family Test	Determine relationship between $E(y\|x)$ and $Var(y\|x)$	GLM	gamma is consistent

heteroskedasticity, unless the data are properly retransformed (Manning, 1998). The presence of heteroskedasticity and the findings of the diagnostic tests indicate that GLM and EEE are more appropriate for these data. However, the GLM Family test (or a modified Park Test) shows that the gamma distribution combined with the log-link function may not appropriately capture the mean-variance relationship and result in the loss of efficiency, even though they are typically used to model healthcare data—this could lead to imprecise results.[5] The extended GLM estimator, or the EEE, developed by Basu and Rathouz (2005), deals with the limitations of GLM in that it does not require choosing the link and variance function a priori and offers a richer set of link and family functions. It helps to avoid bias from choosing a wrong link function and loss of efficiency from having to specify link and family functions separately by solving additional estimating equations. Basu (2005) developed a Stata command to implement this estimator, which is used in this study along with log-OLS (with White heteroskedasticity-robust standard errors) and GLM. By comparing the results of these most-commonly used estimators, this study contributes to the growing literature in health econometrics on model choice when using healthcare data. The importance of getting the variance structure "right" is especially important here for inference. Each of

[5] Manning and Mullahy (2001) discuss that using the quasi-likelihood approach of implementing GLM can protect against some of the efficiency problems arising from misspecification in the distribution function.

these methods is discussed in more detail below.[6]

Finally, these tests are done on positive values of the sample: fewer than 1% of women in the sample had zero non-pregnancy related expenditures. Dropping them from the estimating sample does not cause any systematic bias. However, there are many instances in healthcare data when a large fraction of the sample has zero spending; because of the nature and behavioral aspects of the demand for medical care, this censoring at zero is different from similar situations in other types of data. The zeros may reflect a specific type of behavior and originate from the same or different data generating process as the positive values. This requires careful consideration in modeling approaches. Several of the empirical approaches used to answer the main questions in this study use the subsamples of the data with a nontrivial number of zero values for expenditures and utilization. These approaches must use two-part models, which are also discussed below. Additional empirical methods are described in the relevant sections.

This section focused on issues and tests related to the expenditure data. Deb et al. (2013) mention that most of these issues, such as high zero mass and skewness, are typical of all healthcare data, including utilization, which is considered in this study alongside the cost data. Moreover, the authors highlight that utilization data are usually intrinsically heteroskedastic, with variance increasing with the mean. The nature of the utilization data in the study sample is not examined here since there are fewer debates in the literature on the types of models to use for healthcare count data. Instead, these methods are discussed and compared in the next section.

5.3 Empirical Methods

5.3.1 Least Squares Methods

The OLS model with log-scale dependent variable has been used widely in the estimations of costs in healthcare. The basic regression equation is

$$ln(y) = \mathbf{X}\beta + \varepsilon$$

[6] This study is focused on discovering the effect of nonlinear pricing on the mean outcomes, not median, and hence uses mean-based methods rather than those used to estimate median outcomes.

with assumption of $E(\varepsilon) = 0$ and $E(\mathbf{X}'\varepsilon) = 0$. Given the identification strategy mentioned in Section 5.1, it is reasonable that both assumptions hold here. However, the tests above show that another OLS assumption that may not hold up in these data is the normality of the error term. Doshi et al. (2006) discuss that this violation may not affect regression coefficients, but it does effect tests for statistical significance because it causes problems for retransformation; moreover, the central limit theorem can be called upon in large samples, making this issue somewhat trivial. However, Deb et al. (2013) and Manning and Mullahy (2001) argue that non-normality of errors on the log-scale will cause problems for retransformation of regression coefficients into raw scale because the standard retransformation factor $E(\exp(\varepsilon))$ would produce inconsistent estimates. Additionally, the tests in the previous sections show that this data have complex heteroskedasticity, which violates the OLS assumption of constant variance. So, if ε is heteroskedastic in some or all $\mathbf{X}$, $E(\exp(\varepsilon))$ is some function of $\mathbf{X}$, which gives [7]

$$E(y|\mathbf{X}) = f(\mathbf{X})\exp(\mathbf{X}\beta) \Longrightarrow \ln(E(y|\mathbf{X})) = \mathbf{X}\beta + \ln(f(\mathbf{X}))$$

The form of $f(\mathbf{X})$ is unknown since heteroskedasticity may arise from several sources; hence, it may be complicated to model empirically, which causes problems for retransformation in log-OLS. Manning and Mullahy (2001) discuss that the presence of heteroskedasticity may cause bias in the coefficients and suggest using GLM methods with a quasi-likelihood approach instead. These modes require only the correct specification of the mean function to preserve consistency of the coefficients and test statistics; misspecification of the distribution function may only affect efficiency. However, since log-OLS does not fail other tests and is usually "the first step" in any analysis, the results of the log-OLS model are presented throughout the paper, with caution about using them for interpretation and conclusions.

A special note must be made on obtaining marginal effects after log-OLS since $\ln(E(y|\mathbf{X})) \neq E(\ln(y|\mathbf{X})$:

$$\exp\left(\ln(y)\right) = y = \exp\left(\mathbf{X}\beta + \varepsilon\right) = \exp(\mathbf{X}\beta)\exp(\varepsilon)$$

[7] Adapted from Manning and Mullahy, 2001.

Taking the expectation, the expression becomes:

$$\begin{aligned} E(y|\mathbf{X}) &= E\left[\exp(\mathbf{X}\beta)\exp(\varepsilon)\right] \\ &= \exp(\mathbf{X}\beta)E\left[\exp(\varepsilon)\right] \\ &= \exp(\mathbf{X}\beta)\rho(\mathbf{X}) \end{aligned}$$

where $\rho(\mathbf{X}) = E\left[\exp(\varepsilon)|\mathbf{X}\right]$ is the retransformation factor. The marginal effect in this case is a nonlinear transformation of the regression coefficients that faces two issues. First, there is uncertainty arising from the $\mathbf{X}s$: in the logged scale, the multiplicative effect on the expenditures is estimated holding all else equal, but the exponentiation of predicted log expenditures brings imbalances in the covariates (Doshi et al., 2006). Secondly, there is uncertainty in the retransformation factor itself. Duan (1983) shows that simple exponentiation of the predicted log expenditures leads to biased coefficients. There is a standard retransformation factor that can be applied to the case with homoskedastic and normally distributed error. Additionally, Duan (1983) suggests a smearing estimator to be applied when the errors homoskedastic, but not normal. However, in the case of these data, both assumptions are violated. Deb et al. (2013) suggest using Duan's smearing estimating to correct for non-normality and adapting it for heteroskedasticity by group.

While the tests in Section 5.2 show that these data are likely to have heteroskedasticity from multiple sources, it is reasonable to assume that one of the main sources is in the labor variable. The formal test of this assumption shows that indeed there is heteroskedasticity associated with being in the labor (vs. the non-labor) group because there are inherent differences between the two groups. This subgrouping is used to construct the retransformation factor that is applied to predictions and calculation of marginal effects. The formal expression for this factor is

$$E\left[\exp(\varepsilon)\right] = \frac{1}{N_L}\sum_{i \in g^L} \exp(\varepsilon_i jt)$$

where N_L is the number of women in the labor group, and g^L is whether the women is in the labor group. Then, the expected value of the outcome variable is

$$E[y] = \exp(\mathbf{X}\beta)D^L_{smear}$$

where D^L_{smear} is Duan's smearing estimator by labor group defined as

$$D^L_{smear} = \frac{1}{N_L} \sum_{i=1}^{N} \exp(\varepsilon_i)$$

Even though this method corrects for non-normality and heteroskedasticity, the computation of marginal effects has to account for all sources of uncertainty: both in the estimated coefficients and in the retransformation factor. Hence, standard errors are computed by bootstrapping (Deb et al., 2013).

5.3.2 Generalized Linear Model

The GLM model has been recommended in the literature for healthcare data because it accommodates skewness and related issues via variance-weighting rather than transformation models, which, as discussed above, produce extra challenges (Manning and Mullahy, 2001; Deb et al., 2013). The building blocks of the model are the mean function $\mu(x) = E(y|x)$, with $y \sim F$ (F is distributional family), and the link function $g\left[E(y)\right] = \mathbf{X}\beta$ that relates the mean on the raw scale to the covariates. Since the literature recommends using the log-link function for expenditure data, then[8]

$$\begin{aligned} &g(\mu) = ln(\mu) \\ &\text{where } \mu = \exp(\mathbf{X}\beta) \end{aligned}$$

Here, $g(\cdot)$ is a strictly monotone differentiable link function that relates μ to the linear predictor η. The variance function is

$$Var(y|x) = \alpha \times \left[E(y|x)\right]^{\gamma} = \alpha \times \left[\exp(\mathbf{X}\beta)\right]^{\gamma}$$

GLM model implies a choice of "standard" parametric distributional assumptions. After applying the modified Park test described in Section 5.2, gamma family is chosen because the estimated value of $\gamma \approx 2$ (Manning and Mullahy, 2001). Then, the family-specific variance function is proportional to $E(y|x)^2$:

$$V(y|x) = \mu^2$$

[8] Adapted from McCullagh and Nelder, 1989 and Basu, 2005.

The diagnostic tests showed that the GLM model would perform well in estimations using the given dataset, but may be less efficient than log-OLS, while being more consistent (Deb et al., 2013). Besides, there is no need for retransformation when calculating marginal effects:

$$\frac{\partial \mu}{\partial x_i} = \frac{\partial g(\mu(x_i))}{\partial x_i} = \frac{\partial g(\cdot)}{\partial \mu} \cdot \frac{\partial \mu}{\partial x_i} = \beta_{x_i} \exp(\mathbf{X}\beta)$$

5.3.3 Extended Generalized Linear Model

Deb et al. (2013) discuss that the modified Park test is based on a two-step process that does not correct for the uncertainty from the estimated coefficients and may also be sensitive to right hand side specification. They emphasize the need for a more flexible GLM setup with a richer set of link and family functions. This type of approach was introduced by Basu and Rathouz (2005) and is referred to as the Extended Estimating Equations method. While GLM improves on the log-OLS by dealing better with heteroskedasticity and avoiding problems introduced by retransformation, EEE approach further allows to relax assumptions made about the functional forms of hateroskedastic variance and mean function, which have to be specified *a priori* in GLM model. Moreover, the effect of a covariate on the outcome in log-transformed OLS and GLM log-link are specified as multiplicative. A misspecification of either the mean function or the variance function can lead to bias and inefficiency (Basu, 2005).

The new estimator proposed by Basu and Rathouz (2005) relaxes the need to prespecify the scale of estimation or the functional form of the variance function, and allows the estimation of both the flexible mean function and variance structure using the data. This is done via additional estimating equations to estimate the ancillary parameters that specify the optimal link and variance functions from the data simultaneously with estimating the regression coefficients. Basu (2005) provides a Stata command to implement this estimator. The model description is as follows, using similar notation from Section 5.3.2. The mean function is defined as $\mu(x) = E(y|x)$, and the link function is $g\left[E(y)\right] = \mathbf{X}\beta$. A parametric family of link functions indexed by λ is defined as:

$$g(\mu, \lambda) = \begin{cases} (\mu^{\lambda} - 1)/\lambda, & \text{if } \lambda \neq 0 \\ \ln(\mu), & \text{if } \lambda = 0 \end{cases}$$

Similarly, a family of variance functions is defined as $h(\mu;\theta_1,\theta_2)$ indexed by (θ_1,θ_2). Two types of families can be used in the model. The one used here is power variance family, which is the default:

$$Var(y) = \theta_1\mu\theta^2$$

This variance family includes, as special cases, the variances of several standard distributions using for modeling health outcomes.[9] For example, for Gamma distribution, $\theta_1 > 0$ and $\theta_2 = 2$. The three parameters λ, θ_1, and θ_2 are estimated along with the regression coefficients, which allows to make some conclusions about the optimal mean and variance functions chosen. The marginal effect is calculated as follows:

$$\frac{\partial\mu}{\partial x_i} = \frac{\partial g(\mu(x_i))}{\partial x_i} = \frac{\partial g(\cdot)}{\partial\mu}\cdot\frac{\partial\mu}{\partial x_i} = \mu(1-\lambda)\beta_{x_i} = \left[(\lambda\mathbf{X}\beta+1)^{\frac{1-\lambda}{\lambda}}\right]\beta_x$$

Marginal effects for both the GLM and the EEE models are calculated using the -margins- command in Stata 13.

5.3.4 Two-Part Models

Healthcare data are often characterized by a high zero mass for certain subgroups of the study population or certain types of medical services. In one scenario discussed in Section 7 of this paper (the first trimester analyses), the cost and utilization data have a substantial number of the sample with zero values. In this case, it is reasonable to assume that the zeros are generated by a separate process that the positive values: these zeros are "true" zeros, and not censored values. First, a pregnant woman chooses whether to purchase any non-pregnancy related care. If she decides to use some of this care, she then decides on the amount of care to purchase. Two-part models (2PM) are usually suggested to handle the large zero mass. The first step estimates the probability of having non-zero expenditures or utilization via logit or probit. The second part estimates the level of expenditures or use for $y > 0$. Appropriate continuous or count data models can be used here. Predictions and marginal can be made for the whole sample or for each part separately.[10]

[9] The other family is quadratic variance: $Var(y) = \theta_1\mu + \theta_2\mu^2$. Basu (2005) discusses that for the gamma distribution either power or quadratic variance formulation can be used.

[10] Heckman selection models are also used to estimate two-part models. These approaches are usually used when selection bias is suspected, and the two parts of the decision-making process are connected. In this case, the decision-making process is assumed to be a two-part process, without underlying selectivity.

More formally, the first part of the model can be written as $Pr(y > 0|\mathbf{X}) = \pi(\mathbf{X})$. Assuming the log-transformation for the "level" of y, the second part can be written as [11]

$$\ln(y) = \ln(\mu(\mathbf{X}\beta)) + \varepsilon, \quad y > 0$$

If $\ln(\mu(\mathbf{X}\beta)) = \mathbf{X}\beta$ and $E[\varepsilon|y > 0, \mathbf{X}] = 0$, then

$$E[\ln(y)|y > 0, \mathbf{X}] = \mathbf{X}\beta$$

The full regression equation can constructed using the basic rule of probability.

$$\begin{aligned} E[y|\mathbf{X}] &= Pr(y > 0|\mathbf{X}) \times E(y|y > 0, \mathbf{X}) \\ &= \pi(\mathbf{X}) \times \{\mu(\mathbf{X}) \times E\left[\exp(\varepsilon) \mid y > 0, \mathbf{X}\right]\} \\ &= \pi(\mathbf{X}) \times \{\mu(\mathbf{X}) \times \rho(\mathbf{X})\} \end{aligned}$$

since $y = \mu(\mathbf{X})\exp(\varepsilon)$ for $y > 0$. The notation $\rho(\mathbf{X})$, following Mullahy (1998), is used instead of a constant to highlight that the error retransformation is usually a function of $\mathbf{X}$, and not a constant scalar.

Parametrizing each component:

$$E(y|\mathbf{X}) = \pi(\mathbf{X}\alpha) \times \mu(\mathbf{X}\beta) \times \rho(\mathbf{X}\gamma) \tag{5.2}$$

The first part of the model $Pr(y > 0|\mathbf{X}) = \pi(\mathbf{X})$ is usually estimated using logit or probit. This study uses a logit specification:

$$Pr(y > 0|\mathbf{X}) = \frac{1}{1 + \exp(-\mathbf{X}\alpha)}$$

Deb et al. (2013) propose a Stata command to evaluate a two-part model, via logit (or probit) in the first step, and log-OLS or GLM in the second step. They also propose a way to estimate marginal effects. However, they highlight that the retransformation problems, similar to the ones described in Section 5.3.1, persist here in the presence of heteroskedasticity in the second part of the 2PM, so bootstrapping of standard errors is recommended.

Mullahy (1998), however, argues that the retransformation problem may be much more serious and may impact the consistency of the 2PM point estimates and partial

[11] Adapted from Mullahy, 1998.

effects. Even though $E[\varepsilon|y > 0, \mathbf{X}] = 0$ still holds, it does not mean that $E[\phi(\varepsilon)|y > 0, \mathbf{X}]$ (where $\phi(\cdot)$ is a distribution-robust retransformation factor) is a constant not depending $\mathbf{X}$. In fact, the diagnostic tests showed complex heteroskedasticity in the data used for this study. So, the estimates for $E(y|y > 0, \mathbf{X})$ are obtained in the following way:

$$\begin{aligned} E[y|y > 0, \mathbf{X}] &= E[\exp(\mathbf{X}\beta)\exp(\varepsilon)|y > 0, \mathbf{X}] \\ &= \exp(\mathbf{X}\beta) \times E[\exp(\varepsilon)|y > 0, \mathbf{X}] \\ &= \exp(\mathbf{X}\beta) \times \rho(x) \end{aligned}$$

where $\rho(\mathbf{X})$ is the retransformation factor under the assumption $E[\varepsilon|y > 0, \mathbf{X}] = 0$. Mullahy (1998) says that the component $\rho(\mathbf{X})$ presents problems for parameter estimates and partial effects because ρ may depend on $\mathbf{X}$ in a non-trivial manner.

Mullahy proposes that the exponential conditional mean specification $\rho(\mathbf{X}) = \exp(\mathbf{X}\gamma)$ may be a reasonable parametric model for $\rho(\mathbf{X})$. He then specifies a modified two-part-model (M2PM), the basic idea of which is to replace the second part of homoskedastic 2PM with the assumption that

$$E[y|y > 0, \mathbf{X}] = \exp(\mathbf{X}\beta) = \mu(\mathbf{X})$$

so that

$$y = \exp(\mathbf{X}\beta) \times \exp(\varepsilon), \quad y > 0$$

where $E[\exp(\varepsilon)|y > 0, \mathbf{X}] = 1$. This is essentially an exponential conditional mean (ECM) specification. Using the logit specification from the first part of the model, Mullahy shows that it is possible to recover $E[y|\mathbf{X}]$ from the two parts of M2PM:

$$\begin{aligned} E[y|\mathbf{X}] &= Pr(y > 0)|\mathbf{X}) \times E[y|y > 0, \mathbf{X}] \\ &= \frac{\exp(\mathbf{X}\alpha)\exp(\mathbf{X}\beta)}{1 + \exp(\mathbf{X}\alpha)} \\ &= \frac{\exp(\mathbf{X}(\alpha + \beta))}{1 + \exp(\mathbf{X}\alpha)} \\ &= \Psi(\mathbf{X}; \theta) \end{aligned}$$

where $\theta = [\alpha, \beta]$ is the vector of regression coefficients. This specification allows $E[y|\mathbf{X}]$ to be specified as two components, not three, as in the traditional 2PM (see equation

5.2).[12] Empirically, this can be estimated similarly to the 2PM model. The first part estimated with a logit (or probit) to get α in $Pr(y > 0|\mathbf{X};\alpha)$. The second part is estimated via non-linear least squares methods (NLLS), with the key orthogonality condition needed for identification: $\rho(\mathbf{X}) = E[\exp(\varepsilon)|y > 0, \mathbf{X})$ being a constant not depending on $\mathbf{X}$, using the normalization $\rho(\mathbf{X}) = 1$ since $\mathbf{X}$ contains a constant term. Both traditional 2PMs are used alongside the modified 2PMs in Section 7.2 where the outcome variable is expenditures during the first trimester.

5.3.5 Count Data Models

The nature of utilization data, which is discrete and non-negative, cannot be accomodated by linear regression methods. Count data models are typically nonlinear: $E[\mathbf{X}\beta]$ is usually a single-index function like $\exp(\mathbf{X}\beta)$ (Cameron and Trivedi, 2009). The canonical model for estimating count data models is the Poisson model that is based on the underlying Poisson distribution. The mean is $\mu = \exp(\mathbf{X}\beta)$, the variance is $\sigma^2 = \exp(\mathbf{X}\beta)$, and the probability density function is

$$Pr(Y = y|\mathbf{X}) = \frac{\exp(-\mu)\mu^y}{y!}$$

This specification accounts for nonlinearity, intrinsic heteroskedasiticty, and non-negativity. This model is usually estimated using Maximum Likelihood methods. However, the equality of mean and variance (or equidispersion) is usually an unrealistic assumption for these data. Deb et al. (2013) mention that the quasi-maximum likelihood for a Poisson regression relaxed the assumption of mean-variance equality. This model is used here. There are still concerns about efficiency if the variance function is misspecified (Deb et al., 2013).

Using a method proposed by Cameron and Trivedi (2009), a formal test of the hypothesis of equidispersion is conducted. The null hypothesis is $Var(y|\mathbf{X}) = E(y|\mathbf{X})$. The alternative hypothesis is based on the equation:

$$Var(y|\mathbf{X}) = E(y|\mathbf{X}) + \alpha^2 E(y|\mathbf{X})$$

which is the variance function of the Negative Binomial 2 model. The test is on $H_0: \alpha = 0$ against $H_a: \alpha > 0$. The outcomes shows the presence of significant overdispersion,

[12] For derivation of partial effects based on this specification, see Mullahy (1998).

with variance exceeding the mean. The Negative Binomial model is applied in the case with overdispersion.[13] The mean is formulated the same as in the Poisson model, with the variance function defined as

$$Var(y|\mathbf{X}) = \mu + \alpha g(\mu) > \mu$$

Usually, α is treated as constant, and variance is defined as either linear (Negative Binomial-1) or quadratic (Negative Binomial-2) function. The latter is used in this study:

$$Var(y|\mathbf{X}) = \mu + \alpha\mu^2$$

Moreover, similar to the discussion in Section 5.3.4, the exponential conditional mean (ECM) specification can also used to account for nonlinearity of the utilization data as it assumes a nonlinear relationship for the utilization regression:

$$E(y|\mathbf{X}) = \mu = \phi \exp(\mathbf{X}\beta)$$

The ECM specification accommodates the typically skewed nature of the distribution and implies that the effect of covariates is proportional rather than additive, with a constant proportional effect (Jones, 2010). The ECM model can be viewed as a nonlinear regression and is estimated via non-linear least squares. Mullahy (1997) emphasizes that this specification is better suited to deal with potential unobserved heterogeneity that may arise during the data generating process than the traditional count data models and is more suitable to data on nonnegative outcomes because it is based on a log-link relationship. He applies the ECM model in the context of count data, specifically cigarette smoking behavior, and suggests its advantages for medical utilization data; it is used here for the demand estimations with the number of visits for non-pregnancy related care as the outcome variable. Following Mullahy (1997), Manning and Mullahy (2001), and Manning (1998), it is assumed that $y|\mathbf{X}$ has an ECM structure with a linear index function. This approach is used throughout the paper, with application to both the expenditure and utilization data.

The zero-mass problem for count data is accomodated by a hurdle-model, similar to the two-part model described in Section 5.3.4 for the cost data. The underlying

[13] For the description and formulation of the Negative Binomial distribution $NB(\mu, \alpha)$, see Cameron and Trivedi (2009).

assumption here is the same as for the cost data: zeros are generated by a different process than the positive values. Formally, the model is [14]

$$\begin{aligned} Pr(Y = 0|\mathbf{X}) &= f_1(0|\theta_1) \\ Pr(Y = y > 0|\mathbf{X}) &= \frac{1 - f_1(0|\theta_1)}{1 - f_2(0|\theta_2)} \times f_2(y|\theta_2) \end{aligned}$$

where

$$\begin{aligned} &f_1(\cdot|\theta_1) \text{ is a Logit/Probit Model} \\ &f_2(\cdot|\theta_2) \text{ is a Poission/Negative Binomial Model} \\ &\frac{1}{1 - f_2(0|\theta_2)} \times f_2(y|\theta_2) \text{ is a Truncated Count Density} \end{aligned}$$

The logit model is used for the first part of the model, with the Poisson and Negative Binomial used in the second part. These models are implemented using built-in and user-written commands in Stata 13.

[14] Adapted from Deb et al., 2013.

Chapter 6

Labor as Price Effect in Demand Analyses

6.1 Introduction and Motivation

The first empirical examination of the research questions outlined in Section 5.1 uses labor as a large health event that moves a consumer past the deductible threshold and thus lowers her expected end-of-year price. Thus, this price is not incorporated directly into the demand estimations, but is "proxied" by labor. This innovative approach considers the annual non-pregnancy related expenditures and utilization by pregnant women and is one of the few studies to examine demand in the presence of a predictable large health event that moves a consumer past the deductible threshold. Three different empirical approaches are used to examine whether pregnant women who give birth in a calendar year have higher demand for medical care than pregnant women without birth, to help understand whether healthcare consumers are myopic or forward-looking. The two main questions examined are:

- Do women who give birth in a calendar year have higher demand for medical care not related to pregnancy?
- Which price, current (CP) or expected-end-of-year (EFP), do they use when making consumption decisions?

6.2 Empirical Approach I: Labor as a Proxy for Expected-End-of-Year Price

The first empirical approach employed to answer the two research questions makes the fewest assumptions about the relationship between labor and demand for non-pregnancy related care. It uses labor as a binary proxy for the EFP, so the main assumption, as discussed in Section 5.1, is that the year of labor is exogenous to a pregnant woman. This is the source of identification for this empirical approach. It represents the first attempt to look at the relationship between giving birth within calendar year and the demand for medical care. The main premise is that the cost of labor and delivery will exceed the deductible, lowering a woman's EFP, and thus labor and delivery is a reasonable proxy for that price. The estimation equation is

$$y_{ijt} = \alpha + \beta b_t + \lambda bm_t + \gamma \mathbf{X}_i + \delta \mathbf{Insur}_j + \theta \mathbf{Insur}_j \times b_t + \tau \mathbf{Year}_t + \varepsilon_i \qquad (6.1)$$

where y_{ijt} is the annual demand (expenditures or use) on non-pregnancy related care, b_t is a binary variable (=1 if a woman gave birth in year t, and 0 otherwise), bm_t is months before labor (described in Section 4.3), and $\mathbf{Insur}_j \times b_t$ is a vector of interaction terms of insurance characteristics and labor. It is assumed that $E[\varepsilon|\text{labor}] = 0$. Including interaction terms of labor with the demographic variables accounts for the possible health-related and other unobservable effects of labor—the binary "labor" variable is expected to have only the price effect on the demand. The coefficient β answers the question of whether women who give labor in a calendar year spend more on non-pregnancy related care because of the lower end-of-year price than those who do not give birth until the following year.[1] To answer the second question, whether they use current or expected future price when making consumption decisions, either the CP or the EFP, or both, must be incorporated explicitly in the model. The EFP is proxied for by the labor variable. The variable "size of individual deductible" can be considered a proxy for CP since it is the price of care before the deductible is met. Hence, $\delta_{\text{ind. deductible}}$

[1] It could be argued that non-pregnancy related care is postponable and thus may not respond to lower price. However, since the study is interested in within-year price response, for those reacting to lower EFP, this price should be a significant determinant of demand. Moreover, care received after labor is likely to have higher opportunity costs (due to taking care of a newborn, etc), so it is reasonable to assume that a forward-looking woman responding to lower EFP would not delay non-pregnancy related care.

associated with this variable can be interpreted as the effect of the increase in the CP on the demand for medical care. The interaction terms between insurance characteristics and labor are added to the model to account for the differential effect that insurance characteristics may have on the demand for the labor group. The estimators used for this approach are log-OLS, GLM and EEE.

Table 6.1 presents the results of these estimations. The coefficient on the "labor" variable is positive and significant across all models, indicating that giving birth has a positive effect on the demand for care, as hypothesized, through changing the expected end-of-year price. The interaction terms between labor and the insurance characteristics are of small magnitude and only marginally statistically significant, so they are not expected to effect the demand in a meaningful way, but are included in the calculation of marginal effects. However, the coefficient on the interaction term "labor×age" appears to have a strong negative effect on spending across all three models. This finding indicates that for the labor group higher maternal age may lower the demand for non-pregnancy related care because there may other health effects related to maternal age that are not identified medically during doctor visits (and hence are not picked up by the billing codes). Since it is established, in the medical literature and in common knowledge, that higher maternal age poses potential risks or complications for pregnant woman, it is possible that women in more advanced stages of pregnancy (i.e. those that would deliver within the calendar year examined) would not choose to undergo more discretionary medical services, regardless of the lower price.

The coefficient on "months before labor" variable is negative and significant, highlighting that labor in later months of the year has a negative effect on spending. This finding hints at the possibility that there is lack of response to the EFP among the labor group women. This would be further investigated in Section 6.4. The coefficient on the "size of individual deductible" variable is positive and significant for log-OLS, but negative and not significant for GLM and EEE.[2] If this variable is taken as a proxy for the current price of care, it is likely not a factor in the demand for care overall and for the labor group in particular, as the interaction term "labor×ind.deductible" is not statistically significant.

[2] For reasons given in Section 5.1, the results from log-OLS are expected to be less precise and even biased; thus, the conclusions are drawn mainly from the results of the GLM and EEE models.

Table 6.1: Labor as EFP Proxy Estimation Results: Expenditures

	OLS	GLM	EEE
Labor	0.15662*	0.55599***	0.54500***
	(0.09247)	(0.16206)	(0.16207)
Months before labor	-0.00952***	-0.05119***	-0.05209***
	(0.00351)	(0.00635)	(0.00682)
Age	-0.11927***	-0.18743***	-0.18588***
	(0.01190)	(0.02104)	(0.02266)
Age squared	0.00252***	0.00382***	0.00378***
	(0.00020)	(0.00035)	(0.00039)
Labor×age	-0.01872***	-0.02252***	-0.02205***
	(0.00272)	(0.00479)	(0.00482)
Family size	-0.01093***	-0.00531	-0.00538
	(0.00376)	(0.00634)	(0.00633)
Labor×family size	0.00656	-0.00035	-0.00010
	(0.00547)	(0.00906)	(0.00914)
Relationship to employee	0.05130***	0.03215	0.03202
	(0.01504)	(0.02682)	(0.02694)
Type of employee	0.01871***	0.01540***	0.01542***
	(0.00249)	(0.00427)	(0.00427)
Individual deductible	0.00034***	-0.00008	-0.00009
	(0.00012)	(0.00018)	(0.00018)
Family deductible	-0.00015***	0.00005	0.00006
	(0.00006)	(0.00009)	(0.00009)
Coinsurance rate	-1.29038***	-0.86180***	-0.84518***
	(0.13414)	(0.22519)	(0.22725)
Preventative care coverage	0.03740	-0.08334*	-0.08487*
	(0.02735)	(0.04639)	(0.04658)
Individual OOP max	0.00001	0.00004	0.00004
	(0.00002)	(0.00003)	(0.00003)
Family OOP max	-0.00001	-0.00003**	-0.00003**
	(0.00001)	(0.00002)	(0.00002)
Labor× ind. deductible	0.00023	0.00036	0.00038
	(0.00017)	(0.00027)	(0.00028)
Labor× fam. deductible	-0.00008	-0.00018	-0.00019
	(0.00008)	(0.00013)	(0.00013)
Labor× ind. OOP max	-0.00003	-0.00008*	-0.00008*
	(0.00003)	(0.00004)	(0.00004)
Labor× fam. OOP max	0.00002	0.00006**	0.00006**
	(0.00001)	(0.00002)	(0.00002)
Labor× coinsurance	0.17618	-0.14778	-0.18595
	(0.18195)	(0.31314)	(0.32275)
Constant	7.77602***	9.60370***	2.50429***
	(0.18209)	(0.32996)	(0.34844)
Observations	38807	38807	38807

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The marginal effects, presented in Table 6.2, provide more insight on the findings. These marginal effects are overall marginal effects and are calculated based on the coefficients from the "labor" and "individual deductible" variables as well as all interaction terms. [3] The combined effect is actually negative and significant: women who give birth within a calendar year spend on average about $134 less than their non-labor cohort, based on the results from the GLM and EEE models. While the findings in Table 6.1 show a positive effect of labor on spending, this effect is most likely attenuated and even reversed by the "labor×age" variable: the effect of age on the demand for non-pregnancy care for the labor group is quite significant in magnitude, but is likely due to its influence as a health-related covariate, rather than through the price effect. In fact, the "cross-over" age (i.e. the age after which the effect of labor on demand becomes negative) is 25 years old; this relatively young age highlights the significant impact that age has on the labor group and its demand for medical care. The marginal effect for the size of the deductible variable is quite small across all models and not significant for the GLM and EEE models, indicating that the size of the deductible may not matter when making decisions about consumption of non-pregnancy related care.

The differences between the log-OLS and the GLM/EEE results highlight the importance of choosing the most appropriate estimator for the given data to draw conclusion. In this instance, the log-OLS marginal effect for the labor variable is much larger than GLM and EEE, so choosing the wrong estimator in this case may overstate the effects. Interestingly, the coefficients, marginal effects, and standard errors for the GLM and EEE models are quite similar, which likely points to the correct choice of both the link and the family functions for the GLM estimator for these data (Manning and Mullahy, 2001).

The outcome of the estimations of the effect of labor on the utilization of non-pregnancy related medical services are presented in Table 6.3. These results confirm the findings from the expenditure data: the coefficient on the "labor" variable is positive and significant, while the coefficients on most of the interaction terms are not significant, with the exception of "labor×age", which is negative and significant. Unlike the spending

[3] For log-OLS model, marginal effects are calculated using Duan's smearing estimator by the labor group, as discussed in Section 5.3.1. Deb et al. (2013) discuss that in the presence of complex heteroskedasticity, even careful retransformation may not lead to the computation of correct standard errors, so the standard errors from the log-OLS are bootstrapped.

Table 6.2: Labor as EFP Proxy Marginal Effects: Expenditures

	OLS	GLM	EEE
Labor	-397.10***	-134.67**	-134.36**
	(30.27)	(50.61)	(53.43)
Individual Deductible	0.51***	0.06	0.06
	(0.12)	(0.17)	(0.17)

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

estimations, the effect of the "months before labor" is positive and significant, indicating that if labor happens in later months, pregnant women in the labor group have more non-pregnancy related visits. However, it is unclear whether it is due to the effect of lower price of care or because of having more time to make doctor visits before giving birth. The coefficient on the "individual deductible" variable is positive and significant, though small in magnitude.

The marginal effects shown in Table 6.4 are overall effects, calculated from the coefficients of the variables of interest and their interaction terms. Similarly to the findings for the expenditures, the results show that those who give labor have fewer non-pregnancy related doctor visits, by about 1 visit on average. Even though the overall effect of the deductible size is positive, it is negligible in magnitude. As for the comparison between the empirical methods employed, the Poisson and Negative Binomial models have similar marginal effects (even though the coefficients are slightly different in magnitude), with the ECM model showing a smaller marginal effect of labor on the demand, but not too different. Otherwise, the coefficients and standard errors produced by the ECM are similar to the other two estimators.

6.3 Empirical Approach II: Labor as Instrument for Expected End-of-Year Price

The second empirical approach is based on the fact that p^e (the expected end-of-year price) is endogenous because it is determined simultaneously with the decision to purchase medical care in the present and future periods. Using the identification strategy

Table 6.3: Labor as EFP Proxy Estimation Results: Utilization

	Poisson	Neg. Binomial	ECM
Labor	0.16944**	0.13435**	0.21382***
	(0.06765)	(0.06592)	(0.07360)
Months before labor	0.01057***	0.01133***	0.00947***
	(0.00240)	(0.00237)	(0.00253)
Age	-0.03950***	-0.04231***	-0.03514***
	(0.00908)	(0.00880)	(0.01000)
Age squared	0.00113***	0.00117***	0.00106***
	(0.00015)	(0.00015)	(0.00017)
Labor×age	-0.01384***	-0.01293***	-0.01483***
	(0.00202)	(0.00197)	(0.00219)
Family size	-0.00882***	-0.00794**	-0.00985***
	(0.00336)	(0.00314)	(0.00373)
Labor×family size	0.00690	0.00655	0.00661
	(0.00433)	(0.00413)	(0.00473)
Relationship to employee	-0.00324	-0.00717	0.00260
	(0.01164)	(0.01121)	(0.01288)
Type of employee	0.01137***	0.01124***	0.01158***
	(0.00191)	(0.00184)	(0.00208)
Individual deductible	0.00032***	0.00035***	0.00028***
	(0.00009)	(0.00009)	(0.00010)
Family deductible	-0.00018***	-0.00020***	-0.00016***
	(0.00004)	(0.00004)	(0.00004)
Coinsurance rate	-0.87355***	-0.89193***	-0.85121***
	(0.10462)	(0.10242)	(0.11176)
Preventative care coverage	0.01093	0.00925	0.01430
	(0.02126)	(0.02035)	(0.02314)
Individual OOP max	-0.00001	-0.00002	-0.00000
	(0.00001)	(0.00001)	(0.00002)
Family OOP max	0.00000	0.00001	0.00000
	(0.00001)	(0.00001)	(0.00001)
Labor× ind. deductible	0.00015	0.00016	0.00015
	(0.00012)	(0.00012)	(0.00013)
Labor× fam. deductible	0.00000	0.00000	-0.00000
	(0.00005)	(0.00005)	(0.00005)
Labor× ind. OOP max	-0.00002	-0.00002	-0.00002
	(0.00002)	(0.00002)	(0.00002)
Labor× fam. OOP max	0.00001	0.00001	0.00001
	(0.00001)	(0.00001)	(0.00001)
Labor× coinsurance	0.26107**	0.28846**	0.21506
	(0.13258)	(0.13086)	(0.13937)
Constant	2.25010***	2.31264***	2.15452***
	(0.13889)	(0.13460)	(0.15175)
Observations	38808	38808	38808

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6.4: Labor as EFP Proxy Marginal Effects: Utilization

	Poisson	Neg. Binomial	ECM
Labor	-0.92***	-0.93***	-0.8***
	(0.1139)	(0.1133)	(0.1166)
Individual Deductible	0.002***	0.003***	0.002***
	(0.0005)	(0.0005)	(0.0005)

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

described earlier, the binary "labor" variable can be thought of as an appropriate instrument for the endogenous price. The set-up here is as follows. Let $p^e = EFP$. The main idea behind this approach is that the EFP is lower for women who give birth than for those who do not, and thus should effect the demand for care accordingly. The reduced form demand equation is

$$y_{ijt} = \alpha_d + \beta_d EFP_{it} + \gamma_d \mathbf{X}_i + \tau_d \mathbf{Year}_t + \varepsilon_{ijt} \tag{6.2}$$

where subscript d indicates that the coefficients are from the demand equation. The coefficient β_d is the main effect of interest.

Given the theoretical definition of p^e from Section 3, the EFP is constructed empirically:

$$EFP = \begin{cases} 1 & \text{when} & \text{total OOP} \leq \text{ind. deductible} \\ \text{Coinsurance} & \text{when} & \text{ind.deductible} < \text{total OOP} \leq \text{ind. OOP max} \\ 0 & \text{when} & \text{ind. OOP max} < \text{total OOP} \end{cases}$$

Descriptively, the results of this construction match the underlying assumption for this study: women who give birth within a calendar year have a lower expected end-of-year price than those who do not, because the costs of labor and delivery exceed the size of the deductible with certainty. Indeed, in the study sample, the labor group has a lower EFP than non-labor group on average: $EFP_{labor} = 0.10 < 0.39 = EFP_{non-labor}$. The EFP of the labor group is slightly less than the average coinsurance rate of 0.13, indicating that some women who gave birth may have exceeded their individual OOP maximum, resulting in a zero EFP. Figure 6.1 presents a graphical illustration of this result: the cumulative distribution functions (cdf) show the marginal price of the labor group is lower for the entire distribution.

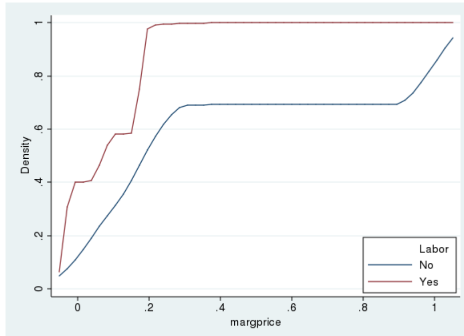

Figure 6.1: Distribution of Expected Marginal Price

With "labor", "months before labor", insurance characteristics, and their interaction terms with labor as determinants of the EMP, the first stage equation is

$$EFP_{ijt} = \alpha + \beta b_t + \lambda bm_t + \gamma \mathbf{X}_i + \delta \mathbf{Insur}_j + \theta \mathbf{Insur}_j \times b_t + \tau \mathbf{Year}_t + \eta_{ijt} \quad (6.3)$$

where $Z = (b_t, bm_t\mathbf{Insur}_j, \mathbf{Insur}_j \times b_t)$. It is reasonable to assume that these instruments fulfill the first requirement for the valid instrumental variables: they are highly correlated with the endogenous variable — $E(EFP_{it}|\mathbf{Z}) \neq 0$. However, given the results from Empricial Approach I in Section 6.2, they likely do not pass the second requirement – $E(\varepsilon_i|\mathbf{Z}) = 0$ – since most of the variables in $\mathbf{Z}$ have a direct effect on the outcome variable (Wooldridge, 2010). Even though "labor" is exogenous to the pregnant women, it is likely to be correlated with the error term in the regression equation (6.2) by influencing the demand for medical care. Hence, additional identification variables are needed for this approach to work.

Insurance characteristics other than the size of individual deductible deductible, coinsurance rate, and individual out-of-pocket maximum, which are used to construct EFP as discussed below, could be used as instruments for the EFP. These variables such as family-level deductible and and family-level out-of-pocket maximum, capture the "generosity" of an insurance plan. However, given the results in the previous section, these variables may also be correlated with the error term. Moreover, they would be zero for a single-enrollee plan. What is more likely to be strictly exogenous are other variables available in the data that describe the plan generosity and influence EFP, but without a direct effect on the demand for care. The plan provisions for out-of-network benefits are not expected to impact the demand for care directly, since examination of similar data shows that most plan enrollees use in-network providers (Kowalski, 2009). Additionally, copayment amounts for various types of services are good indicators of plan generosity and likely impact the expected end-of-year price. Due to limitations of the data, the information on copayments was available for most years for the following types of services: primary care, specialists, inpatient care, and emergency room. Note that copayment amounts may or may not count towards the amount of the deductible, and there is no information on this in the data. Preventative care coverage is also used as an instrument and an indicator of plan generosity since non-pregnancy related care is most likely not part of preventative care. Therefore, the first stage equation formally remains the same as (6.3), but the vector of insurance characteristics, **Insur**, now includes the following variables: individual and family-level deductible for out-of-network providers, coinsurance rate for out-of-network providers, individual and out-of-pocket maximum for out-of-network providers, preventative care coverage, copayment for primary care, specialists, inpatient care, and emergency room services. Interaction terms for the insurance characteristics and labor are also included.

Table 6.5 presents the results of the first-stage equation: the determinants of the EFP as noted in the equation (6.3). Labor indeed has a negative and significant effect on the EFP, confirming the descriptive findings in Figure 6.1. As hypothesized, women who give birth in a calendar year face a lower expected end-of-year price. "Months before labor" is positive and significant, which is intuitive: women who give birth later in the year have more months to pay a higher price for care before reaching the deductible. Most of the insurance characteristics chosen as instruments are statistically significant, which makes them reasonable determinants of the EFP.

The findings from Approach I show that labor has a direct effect on the demand for non-pregnancy related care, and is therefore a determinant of this spending. These results show that while labor may in fact increase the demand for non-pregnancy related care through its price-lowering effect (confirmed here with the lower average EFP for the labor group), the overall effect is negative most likely because of the health considerations associated with birth. Thus, the "labor" variable has to be included in the demand equation in the second stage of the IV estimations as a covariate to control for the potential health effects that labor can have on demand for medical care. Including this variable in both stages allows it to have an impact on demand both through lowering the

Table 6.5: First Stage Equation: Determinants of EFP

	Expected Future Price
Labor	-0.36496***
	(0.01891)
Months before labor	0.00085***
	(0.00020)
Age	-0.01439***
	(0.00292)
Age squared	0.00015***
	(0.00005)
Labor×age	0.00453***
	(0.00053)
Family size	0.00711***
	(0.00114)
Labor×family size	-0.00912***
	(0.00115)
Relationship to employee	-0.01661***
	(0.00346)
Type of employee	-0.00146**
	(0.00063)
Individual deductible–OON	0.00012***
	(0.00002)
Family deductible–OON	0.00000
	(0.00001)
Coinsurance rate OON	0.00006
	(0.00014)
Preventative care coverage	0.00132
	(0.00660)
Individual out-of-pocket max OON	-0.00001***
	(0.00000)
Family out-of-pocket max OON	0.00001***
	(0.00000)
Copayment: primary care	-0.00855***
	(0.00071)
Copayment: specialists	-0.00161***
	(0.00048)
Copayment: emergency room	0.00105***
	(0.00009)
Copayment: inpatient care	-0.00002
	(0.00007)
Labor×ind. deductible OON	-0.00012***
	(0.00002)
Labor×fam. deductible OON	0.00001
	(0.00001)
Labor×ind. OOP max OON	0.00001***
	(0.00000)
Labor×fam. OOP max OON	-0.00001***
	(0.00000)
Labor×coinsurance OOP	0.00088***
	(0.00014)
Labor×copay primary care	0.00347***
	(0.00070)
Labor×copay specialists	0.00119**
	(0.00048)
Labor×copay ER	-0.00039***
	(0.00008)
Labor×copay inpatient	0.00000
	(0.00007)
Constant	0.63654***
	(0.04625)
Observations	36354

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

EFP and via its effect as a particular health condition (i.e. as a direct determinant of spending). The "months before labor" variable is also included to control for potential health heterogeneity due to the time in the gestation period. The new reduced form equation is

$$y_{ijt} = \alpha_d + \beta_d EFP_{it} + \theta_d b_t + \lambda_d bm_t + \gamma_d \mathbf{X}_i + \tau_d \mathbf{Year}_t + \varepsilon_{dijt} \tag{6.4}$$

To implement this approach empirically, several instrumental variables (IV) methods are used. The traditional two-step least squares (2SLS) IV method may not be best suited for this data as it assumes homoskedasticity and thus may cause efficiency problems with the heteroskedastic data in this sample (Cameron and Trivedi, 2009). Thus, the first method applied is a widely used estimator applied in a variety of empirical settings: a generalized method of moments (IV-GMM) estimator, formally defined as

$$\hat{\beta} = (\mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'y$$

where $\mathbf{W}$ is a full-rank symmetric-weighting matrix (Cameron and Trivedi, 2009). Wooldridge (2010) highlights that the choice of the weighting matrix is important for efficiency, and the optimal choice of $\mathbf{W}$ that leads to the smallest variance is defined as

$$\hat{\mathbf{W}} = \left(N^{-1} \sum_{i=1}^{N} \mathbf{Z}'\varepsilon\varepsilon'\mathbf{Z} \right)^{-1}$$

This estimated weighting matrix allows for heteroskedasticity and leads to consistent estimations of $\hat{\beta}$, which is appropriate for these data.

However, Mullahy (1997) suggests that the traditional IV methods may lead to inconsistent estimations due to the assumption of additive separability in the unobserved components, which makes it necessary to make strong assumptions about the statistical relationships between the instrumental variables and the endogenous covariates and the error term in the reduced-form equation, especially in nonlinear models. He proposes using a transformed IV model based on the ECM specification that could be applied to both continuous and count nonnegative-valued data. The conditional mean is written as [4]

$$E[y|\mathbf{X}, \varepsilon] = \exp(\mathbf{X}\beta + \varepsilon)$$

[4] Adapted from Mullahy (1997).

This structure implies a regression model of the form

$$y = \exp(\mathbf{X}\beta)\eta + \tau$$

where $\eta = \exp(\varepsilon)$ is the interpersonal heterogeneity component. Note that $\mathbf{X}$ includes the endogenous variable EFP. Since a constant term is included in $\mathbf{X}$, $E(\eta) = 1$. The regression error τ has the form $E(\tau|\mathbf{X}, \varepsilon) = 0$. This is an exponential regression model for nonnegative dependent variables with multiplicative unobserved heterogeneity with the additional assumption that some regressors are endogenous. Since $\mathbf{X}$ contains an endogenous regressor (i.e. $E(\eta|\mathbf{X}) \neq E(\eta) = 1$), there is a vector of instruments $\mathbf{Z}$ such that $E(\tau|\mathbf{X}, \mathbf{Z}) = 0$ and $E(\eta|\mathbf{Z}) = 1$. These satisfy the standard assumptions needed for a consistent IV estimator.

Mullahy (1997) shows that using a standard residual function as a conditional moment restriction in the GMM estimation of an ECM model with multiplicative unobserved heterogeneity will not yield consistent parameter estimates, because η is not additively separable from the observables in the standard residual function. He proposes a transformation of the regression model to obtain a residual function in which η is additively separable from the potential endogenous regressors. This transformed residual is

$$(\eta - 1) + \exp(-\mathbf{X}\beta)\tau$$

so that the consistent GMM can be based on the conditional moment restriction

$$E[(\eta - 1)|\mathbf{Z}] = E[\exp(-\mathbf{X}\beta)y - 1|\mathbf{Z}] = 0$$

This is a semiparametric nonlinear instrumental variable estimator. In this study it is used for both the expenditure and utilization data and is referred to as "Mullahy GMM." Additionally, for the utilization data an instrumental variable Poisson model is used. It is implemented via -ivpois- command in Stata 13 written by Nichols (2007). It uses a GMM estimator of Poisson regression that allows additional exogenous variables that have not direct impact on the dependent variable, and endogenous variables to be instrumented by the excluded instruments. This procedure is based on Mullahy GMM, but does not impose an ECM structure on the data and includes a multiplicative error term (Nichols, 2007).

Table 6.6 presents the results for the expenditure data for equation (6.4).[5] The first column has the results from the IV-GMM method, and the second shows the results from Mullahy's GMM. The coefficient on the "EFP" variable is negative and significant across both methods, indicating that pregnant women respond to the EFP when making purchasing decisions. However, the findings for the "labor" variable are quite different. The IV-GMM approach shows that labor has a negative significant effect on the demand for non-pregnancy related, while Mullahy-GMM method shows a positive significant effect, similar to the findings from Approach I. This latter finding indicates that the strongest impact of labor on spending comes through its price lowering effect, but it may also influence demand as a particular health condition. The "months before labor" variable is negative and significant across both methods, echoing the findings about its impact on spending from Approach I. Other coefficients are similar across the methods. Mullahy (1997) makes a convincing case of why using the ECM structure for modeling healthcare data leads to more consistent and efficient results than the traditional IV methods. There is some worry that using the IV-GMM method that is based on standard Least Squares methods and a log dependent variable may produce inaccurate results with these data, as seen in Approach I. Hence, only the Mullahy-GMM results are used to draw conclusions and calculate price elasticities.

Table 6.7 presents the results for the utilization variable using the IV-Poisson and the Mullahy GMM methods. Similar to expenditures, the expected future price has a negative and significant effect on the demand for non-pregnancy care. The "labor" is positive and significant, similar to findings in Approach I for utilization and expenditures (and the findings above for the Mullahy-GMM model and expenditures). The "months of labor" variable is positive and significant, also similar to the findings under Approach I. Once again, the number of visits here may be more reflective of other factors, such as availability for doctor's appointments before giving birth. Overall, both the spending and use estimations show that the expected future price has a negative significant effect on the demand of non-pregnancy related care while labor has a positive impact.

The elasticities obtained from the estimations under Approach II are presented in Table 6.8. The elasticity for the expenditure estimations is calculated using Mullahy's

[5] The sample size is slightly smaller in these estimations because the data for some years includes only a few observations, which poses computational problems for the convergence of Mullahy GMM.

Table 6.6: Labor as Instrument for EFP: Expenditures

	IV-GMM	Mullahy GMM
EFP	-1.15916***	-0.81222***
	(0.09216)	(0.15574)
Labor	-0.23023**	0.33956**
	(0.09025)	(0.15306)
Months before labor	-0.01569***	-0.04798***
	(0.00361)	(0.00594)
Age	-0.12654***	-0.17683***
	(0.01188)	(0.02022)
Age squared	0.00258***	0.00356***
	(0.00020)	(0.00034)
Labor×age	-0.01206***	-0.01800***
	(0.00271)	(0.00476)
Family size	0.00657*	0.00877
	(0.00367)	(0.00578)
Labor×family size	-0.00658	-0.01216
	(0.00553)	(0.00875)
Relationship to employee	-0.01956	-0.01446
	(0.01510)	(0.02586)
Type of employee	0.01165***	0.01026**
	(0.00261)	(0.00420)
Constant	8.12175***	9.36443***
	(0.18986)	(0.31026)
Observations	36354	36354

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6.7: Labor as Instrument for EFP: Utilization

	IV Poisson	Mullahy GMM
EFP	-0.12974**	-0.35803***
	(0.06521)	(0.07862)
Labor	0.01038***	0.13448**
	(0.00242)	(0.06480)
Months before labor	0.01038***	0.00967***
	(0.00242)	(0.00241)
Age	-0.04030***	-0.03730***
	(0.00886)	(0.00883)
Age squared	0.00112***	0.00106***
	(0.00015)	(0.00015)
Labor×age	-0.01141***	-0.01125***
	(0.00199)	(0.00198)
Family size	0.00492	-0.00298
	(0.00407)	(0.00298)
Labor×family size	-0.00366	0.00405
	(0.00298)	(0.00404)
Relationship to employee	-0.03036***	-0.02772**
	(0.01111)	(0.01108)
Type of employee	0.01273***	0.01426***
	(0.00192)	(0.00191)
Constant	2.21895***	2.14093***
	(0.14313)	(0.14198)
Observations	36354	36354

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

GMM approach. The price elasticity for expenditures is about -0.21 and range from -0.12 to -0.28. The elasticities for the utilization variable are calculated using both the IV-Poisson and Mullahy's GMM since the results obtained from both are very similar. The elasticities for utilization are lower: -0.09 – -0.10 (ranging from -0.05 to -0.13). These are similar to the findings from previous studies that consider deductibles in demand estimations (Marsh, 2013; Van Vliet, 2001), but are lower than found by Aron-Dine et al. (2012).

Table 6.8: Price Elasticities

	Price elasticities
Expenditures	-0.21
	(-0.12 – -0.28)
Utilization	-0.09 – -0.10
	(-0.05– -0.13)

Ranges (min-max in magnitude) in parentheses

6.4 Empirical Approach III: Month of Labor as Proxy for Anticipatory Effects

The results from the two previous empirical approaches provide some insights on the first question posed at the end of Section 5.1. While giving birth lowers the EFP for the labor group and has a positive effect on the demand for non-pregnancy care, when combined with other factors, especially age, women in the labor group on average spend and use less of this care than their peers in the non-labor group. Approach II also shows that the women respond to the EPF when making purchasing decisions, while Approach I shows that the current price of care, proxied by the size of the individual deductible, has little to no impact on demand. These findings all point to the presence of some degree of forward-looking behavior. However, the findings on the impact of "months before labor" variable are somewhat puzzling. If indeed the women in the sample were fully forward-looking, this coefficient would not be statistically significant as the women would be aware of their low end-of-year price regardless of the actual month of labor. On the other hand, it may be capturing the health effect of labor: it is reasonable to assume that women later in the gestation period may have a different demand for care than those earlier in their pregnancy.

Hence, additional evidence is needed on whether these women are myopic or forward-looking. One way to examine this further is to look closer at the month of labor, which is generally exogenous to the women (similar to the year of labor). What makes it different from the binary "labor" variable is that it captures the number of months that women have the opportunity to be forward-looking in anticipation of the price change created by labor, so it is similar to the "months before labor" variable. If these women are forward-looking, the month of labor should not matter: the expected end-of-year price will be low, so demand for care should not be affected. However, if they are myopic, then the month of labor would have an impact on their spending decisions because they would be guided by their current price. In other words, the month of labor is a different kind of proxy for the EFP than the binary labor variable: it not only captures the price change, but the timing of the price change. This approach is similar to the one employed by Aron-Dine et al. (2012) where they use the month of joining a health plan to test for myopic vs. forward-looking behavior.

For this estimation, the sample is limited only to the labor group, which eliminates most of the observed and unobserved health heterogeneity that may have been present in the first two approaches when comparing labor and non-labor women. Figure 6.2 shows the distribution of the month of labor among the women in the sample. It is fairly uniform, with mean of 6.3 and standard deviation of 3.34. Hence, no additional selection of the month of birth is suspected, so the identification strategy based on the exogeneity of the timing of labor is likely to hold in this case.

The estimating equation becomes:

$$y_{ijt} = \alpha + \beta Month_{it} + \gamma \mathbf{X}_i + \delta \mathbf{Insur}_j + \tau \mathbf{Year}_t + \varepsilon_{ijt} \tag{6.5}$$

where $Month_{it}$ is month of labor and the rest of the variables are defined as before. The coefficients of interest here are β and $\delta_{\text{ind.deductible}}$. Tables 6.9 and 6.10 present the results and marginal effects for the expenditure variable.[6] The month of labor has a negative and significant impact on the demand for non-pregnancy related care across

[6] The marginal effects for log-OLS are calculated using Duan's retransformation factor to correct for error non-normality (Duan, 1983) and are bootstrapped.

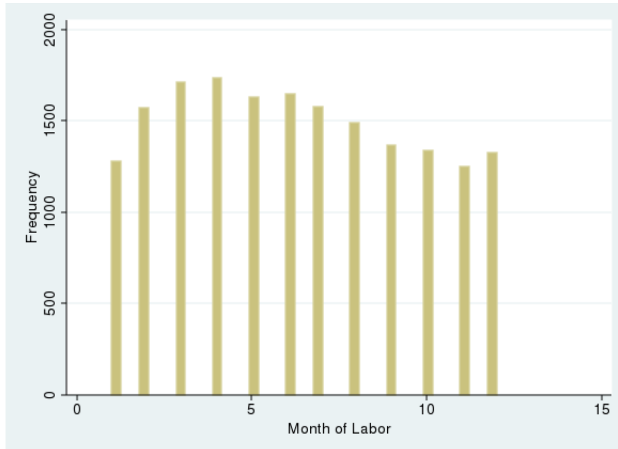

Figure 6.2: Labor Month Frequency Distribution

all models. This finding indicates that when labor happens later in the year, women spend less than when labor happens earlier in the year. This provides some evidence for myopic behavior: since their current price is higher before labor, their demand is lower, on average by about $41, using the estimates from the more trustworthy GLM and EEE models. The size of the deductible does not appear to have a strong effect, even though it is positive, further indicating that the current price may be somewhat of a factor in determining demand for this group. This is not surprising since all of the women in this sample (the labor group) will exceed the deductible with certainty at some point during the year after giving birth.

The results for the utilization variable are different: the effect of an extra month until labor is positive and significant but very small in magnitude (see Tables 6.11 and 6.12), similar to the coefficients on the "months before labor" in Approaches I and II. The fact that the coefficient on labor month is statistically significant indicates that the women may respond to the current price, and hence confirms the presence of myopic behavior. The positive signs on the coefficients across all three models mean that the later the month of labor is, the more visits to the doctor the women make. This can have a different behavioral interpretation, similar to the one given in Section 6.2 for the "labor as IV for EFP" estimates for the utilization outcome: women who give birth in later months of the year have more opportunities for physician visits that those who give birth earlier in the year, for different reasons unrelated to price of care (for example, taking

Table 6.9: Month of Labor: Expenditures

	OLS	GLM	EEE
Month of labor	-0.00741**	-0.04390***	-0.04380***
	(0.00314)	(0.00604)	(0.00724)
Age	-0.04555**	-0.07617*	-0.07714*
	(0.01962)	(0.04080)	(0.04077)
Age squared	0.00097***	0.00157**	0.00159**
	(0.00033)	(0.00071)	(0.00071)
Family size	-0.00355	-0.00355	-0.00370
	(0.00429)	(0.00732)	(0.00730)
Relationship to employee	0.05458**	0.07134*	0.07027*
	(0.02253)	(0.04241)	(0.04227)
Type of employee	0.01929***	0.01836***	0.01830***
	(0.00374)	(0.00635)	(0.00645)
Individual deductible	0.00063***	0.00022	0.00024
	(0.00017)	(0.00023)	(0.00023)
Family deductible	-0.00026***	-0.00011	-0.00012
	(0.00007)	(0.00010)	(0.00010)
Coinsurance	-0.97669***	-0.82836***	-0.82646***
	(0.16976)	(0.27540)	(0.27612)
Preventative care coverage	-0.01965	-0.13331*	-0.13295*
	(0.04230)	(0.07490)	(0.07499)
Individual OOP max	-0.00004	-0.00004	-0.00004
	(0.00002)	(0.00003)	(0.00004)
Family OOP max	0.00002	0.00002	0.00002
	(0.00001)	(0.00002)	(0.00002)
Constant	6.62388***	8.24526***	1.17687**
	(0.29704)	(0.59947)	(0.59837)
Observations	17944	17944	17944

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6.10: Month of Labor: Expenditure Marginal Effects

	OLS	GLM	EEE
Labor	-7.00***	-41.40***	-41.27 ***
	(2.89)	(5.92)	(6.65)
Individual Deductible	0.59***	0.21	0.28
	(0.16)	(0.22)	(0.28)

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

care of a newborn after labor, etc). Hence, the utilization variables for the labor group may reflect behavior unrelated to price. As for the effect of the size of the deductible, it appears to have a slightly stronger impact on utilization than on expenditures: the coefficient is positive and significant across all models, but very small in magnitude.

6.5 Summary

The two main questions posed in Section 5.1 are about the demand for non-pregnancy related care among pregnant women, some of whom give birth in a given calendar year, while others do not, and about which price, current or expected end-of-year, they use when making purchasing decisions. The answer to the first question is revealed by the marginal effects for both spending and utilization: the overall effects show that women in the labor group spend and use less non-pregnancy related care than their non-labor peers, but this is likely to be driven by the health effect of labor rather than the price effect. As for which price they use when making consumption decisions, the findings from Approach I (positive and significant coefficient on the "labor" variable) and Approach II (negative and significant coefficient on the "expected marginal price" variable) provide evidence that these women respond to their expected end-of-year price, which points to the presence of a certain degree of forward-looking behavior. Moreover, the "individual deductible" effect is very small for most of the estimations presented above and mostly not statistically significant, with a few exceptions. If taken as a proxy for the current price, it would appear that this price does not affect purchasing decisions, rejecting the full hypothesis of full myopia and further indicating that this group of consumers is forward-looking.

Table 6.11: Month of Labor: Utilization

	Poisson	Neg. Binomial	ECM
Month of labor	0.00925***	0.00997***	0.00814***
	(0.00215)	(0.00213)	(0.00225)
Age	-0.00477	-0.00422	-0.00659
	(0.01440)	(0.01387)	(0.01533)
Age squared	0.00032	0.00031	0.00035
	(0.00024)	(0.00024)	(0.00026)
Family size	-0.00166	-0.00115	-0.00296
	(0.00300)	(0.00291)	(0.00314)
Relationship to employee	-0.00204	-0.00521	0.00248
	(0.01533)	(0.01498)	(0.01606)
Type of employee	0.01080***	0.01095***	0.01064***
	(0.00256)	(0.00251)	(0.00262)
Individual deductible	0.00049***	0.00053***	0.00043***
	(0.00010)	(0.00010)	(0.00010)
Family deductible	-0.00019***	-0.00021***	-0.00016***
	(0.00004)	(0.00004)	(0.00004)
Coinsurance	-0.53439***	-0.53394***	-0.54121***
	(0.12334)	(0.12025)	(0.12533)
Preventative care coverage	-0.01286	-0.01557	-0.00967
	(0.03014)	(0.02895)	(0.03066)
Individual OOP max	-0.00004***	-0.00005***	-0.00003**
	(0.00002)	(0.00001)	(0.00002)
Family OOP max	0.00002**	0.00002***	0.00002**
	(0.00001)	(0.00001)	(0.00001)
Constant	1.71819***	1.71080***	1.74598***
	(0.21201)	(0.20471)	(0.22482)
Observations	17944	17944	17944

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6.12: Month of Labor: Utilization Marginal Effects

	Poisson	Neg. Binomial	ECM
Month of labor	0.06***	0.07***	0.05 ***
	(0.014)	(0.014)	(0.015)
Individual Deductible	0.003***	0.004***	0.003***
	(0.0007)	(0.0007)	(0.0007)

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

However, the findings from Approach III about the effect of the month of labor, in addition to the coefficients on the "months before labor" variable in Approaches I and II, provide evidence that there is some myopic behavior as care is delayed with the month of labor. These monthly effects need to be examined further to be able to draw appropriate conclusions. Moreover, since the objective of the paper is to provide evidence on myopic versus forward-looking behavior, annualizing healthcare data, spending and utilization, may potentially overstate the time period that a woman has to be forward-looking. For example, those who deliver in the first quarter of the year spend their deductibles and face zero out-of-pocket price for the rest of the year (between 10 and 12 months), so they should use medical care as if it were free for most of the year. It would not matter whether they are forward looking or not. On the other hand, those who deliver in the fourth quarter of the year got pregnant in the first few months of the calendar year and may not have had the reason to be forward-looking because they were not aware of the pregnancy and upcoming labor for some of the initial period. Hence, the time period for "true" forward-looking behavior may be less than a full calendar year (or the insurance benefit period). The section below presents some sensitivity analyses for a sample of pregnant women for which the labor group includes only those who gave birth between April and September of each calendar year.

6.6 Sensitivity Analyses: April-September Births

Annual expenditures and utilization reflect demand for medical care at any time during the year. However, as briefly discussed above, the entire calendar year may not capture the true period of time that a woman in the labor group may have for forward-looking behavior. Those who give birth earlier in the year may have a higher demand because their price is lowered in the beginning of the benefit period and they face that lower price for most of the year; their expected-end-of-year price is essentially their current price for most of the year. On the other hand, women who give birth later in the year may not know their EFP until a few months into the calendar year so their decisions about medical care for some part of the benefit period may be guided by their current price because they may not anticipate a large medical event that would lower this price.

To isolate the true forward-looking behavior period these potentially noisy times, the labor group sample is limited to those who gave birth between April and September of a calendar year. This period is chosen for the following reasons. Women who give birth in the first few months of the year (January-March) face a lower price for care for most of the year. Their interval for forward-looking behavior is quite small because their current price becomes their expected end-of-year price very quickly in the calendar year. On the other hand, those who give birth in the fourth quarter of the year would have became pregnant in the first few months of the year. These women may not know that they are pregnant and will have a lower price for care until later in the year thus shortening their interval for forward-looking behavior. Using only April-September births in the labor group assures that all women in the labor group have similar opportunities for forward-looking behavior. The comparison group in these analyses is all non-labor women in the sample.

The expenditure variable is used as the outcome of interest in this exercise; based on the discussion above, the utilization variable may capture other drivers of demand such as time costs that may not capture the true price response. The same three empirical approaches used for the full sample above are applied, with the GLM model used for Approach I and Approach III, and Mullahy GMM for Approach II.

Table 6.13 shows the summary statistics for the smaller sample. About 30% of the pregnant women in this sample give birth in the April-September window. These women face a much lower expected marginal price. While their total medical spending is much higher than for women in the non-labor group, their non-pregnancy related spending is lower, similar to the descriptive results for the full sample in Table 4.2. Average month of labor is June, which is similar to the month of labor in the main sample.

Table 6.14 presents the results of the estimations for all three approaches, with

Table 6.13: Sensitivity Analyses: Descriptive Statistics

	All Sample	No-labor	Labor
Labor	0.31	0.00	1.00
	(0.46)	(0.00)	(0.00)
Month of labor	-	-	6.38
	-	-	(1.69)
Before labor	-	-	5.38
	-	-	(1.69)
Marginal price	0.30	0.39	0.10
	(0.37)	(0.42)	(0.09)
Age	29.37	29.27	29.60
	(5.54)	(5.73)	(5.09)
Family size	4.17	4.15	4.21
	(2.56)	(2.57)	(2.53)
Relationship to employee	0.53	0.55	0.47
	(0.50)	(0.50)	(0.50)
Individual deductible	365.65	372.67	350.17
	(249.73)	(271.87)	(191.21)
Family deductible	827.43	840.53	798.56
	(553.20)	(590.96)	(457.79)
Coinsurance	0.13	0.13	0.13
	(0.08)	(0.08)	(0.08)
Preventative care coverage	0.85	0.85	0.86
	(0.36)	(0.36)	(0.35)
Ind. OOP max	1871.75	1887.05	1838.01
	(1102.63)	(1111.28)	(1082.60)
Family OOP max	3599.02	3626.88	3537.62
	(2184.16)	(2194.91)	(2159.11)
Total medical spending ($)	4051.26	2026.80	8514.80
	(5397.38)	(3834.63)	(5655.39)
Total non-pregnancy spending ($)	1240.47	1404.15	879.59
	(3283.46)	(3645.63)	(2249.37)
Observations	30327	20864	9463

Note: Standard deviations in parentheses.

Approach II having two types of results, similarly to what was used for the full sample. The first column shows the findings for the "labor as a proxy" for the EFP approach. The coefficient on the labor variable is positive and significant. The marginal effect of labor on the demand for non-pregnancy care, displayed in Table 6.15, show a larger negative significant effect of about $224. These findings are similar to the ones for the full sample under Approach I. While the "labor" variable has a positive and significant impact on spending, when combined with other factors, most notably age, the marginal effect is negative and of a larger magnitude than for the full sample. The results from Approach III are in the second column of Table 6.14 and Table 6.15. The marginal effect is almost identical to the results obtained from the estimation with the full sample: an extra month before labor results in about $40 fewer spending, on average. For both approaches, the size of the deductible has a positive but not significant effect, echoing the findings from the full sample that it likely is not a meaningful determinant of demand for non-pregnancy care.

The third column of Table 6.14 show the results of the estimations from Approach II ("labor as instrument for EFP"). The EFP is negative and significant, with the price elasticity of -0.23 (range between -0.13 – -0.32), which is in line with the results obtained earlier. Labor as a covariate is positive and significant, and the coefficients on the "months before labor" variable for both Approach I and Approach II are negative and significant, just as in the full sample. Overall, the results of the sensitivity analyses mirror the results of the main estimations, confirming the findings.

6.7 Discussion

This study is one of the few to explicitly incorporate nonlinearity in the pricing structure into the estimation of demand for medical care and provide insights on the anticipatory effects in the demand for medical care. This nonlinearity is created by the presence of deductibles in health insurance policies and must be taken into account as the price for care is not constant throughout the year. The concept of the expected end-of-year price is applied to a particular health setting: normally pregnant women who give birth in different periods, and thus face a different expected end-of-year price for medical care. This particular setting presents a new approach to considering how nonlinear pricing

Table 6.14: Sensitivity Analyses Results: Expenditures

	Labor as Proxy	Labor Month	Labor as IV
Labor	0.59273***		0.45105**
	(0.20122)		(0.19107)
Before labor	-0.04297***		-0.05128***
	(0.01453)		(0.01452)
Marginal price			-0.76356***
			(0.16495)
Month of labor		-0.04434***	
		(0.01454)	
Age	-0.20544***	-0.03968	-0.19668***
	(0.02139)	(0.04418)	(0.02050)
Age squared	0.00412***	0.00094	0.00390***
	(0.00036)	(0.00074)	(0.00034)
Labor×age	-0.02411***		-0.02175***
	(0.00523)		(0.00540)
Family size	-0.00641	-0.00516	0.00815
	(0.00642)	(0.00932)	(0.00584)
Labor×family size	-0.00087		-0.01617
	(0.01068)		(0.01047)
Relationship to employee	0.00788	0.01730	-0.02430
	(0.02982)	(0.05772)	(0.02859)
Type of employee	0.01497***	0.02101**	0.01081**
	(0.00488)	(0.00874)	(0.00488)
Individual deductible	-0.00003	0.00015	
	(0.00019)	(0.00030)	
Family deductible	0.00003	-0.00007	
	(0.00009)	(0.00014)	
Coinsurance	-0.90786***	-1.18500***	
	(0.23078)	(0.39524)	
Preventative care coverage	-0.07338	-0.14117	
	(0.05238)	(0.10598)	
Individual OOP max	0.00004	-0.00008*	
	(0.00003)	(0.00004)	
Family OOP max	-0.00003**	0.00003	
	(0.00002)	(0.00002)	
Labor× ind. deductible	0.00021		
	(0.00034)		
Labor× fam. deductible	-0.00011		
	(0.00016)		
Labor× ind. OOP max	-0.00011**		
	(0.00005)		
Labor× fam. OOP max	0.00007**		
	(0.00003)		
Labor× coinsurance	-0.46501		
	(0.40307)		
Constant	9.87678***	7.79531***	9.64859***
	(0.33305)	(0.66926)	(0.31689)
Observations	30327	9463	28345

Standard errors in parentheses

Table 6.15: Sensitivity Analyses: Marginal Effects

	Labor	Labor Month
Labor/month of labor	-224.69**	-40.27***
	(108.72)	(13.07)
Individual Deductible	0.8***	0.97***
	(0.19)	(0.24)

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

affects demand by avoiding the uncertainty in medical care spending (versus Aron-Dine et al., 2012), using a large health event as a way to "cross" the nonlinearity (versus Marsh, 2013), and exploring the predictability of the expenditures (versus Kowalski, 2009) to investigate the presence of the anticipatory effects.

The two research questions posed in this part of the study are whether pregnant women who give birth within a calendar year have higher demand for medical care than those who do not and which price, current or expected future price, they use when making consumption decisions to make conclusions about myopic vs. forward-looking behavior. This empirical section looks at the annual demand, expressed as both expenditures and utilization, for non-pregnancy related care, as type of care that would be discretionary for this sample and hence responsive to price effects. The two research questions are examined via three empirical approaches: labor as a proxy for the EFP, labor as an instrument for EFP, and labor month as a proxy for EFP. The first two approaches use the sample of both labor and non-labor group, while the third focuses only on women who give birth within the calendar year. The annual demand investigation is supplemented by some sensitivity analyses, where the labor group is limited to births within the April-September period in attempts to isolate the period that a woman in the labor group would have for forward-looking behavior.

The findings of the empirical estimations show that labor does have a positive impact on the demand for non-pregnancy related care, both in expenditures and utilization, by lowering the expected end-of-year price. This price is a significant determinant of spending and physician visits, with the elasticities of about -0.10 to -0.21. These findings point to forward-looking behavior in demand for this type of medical care. Additionally, the null hypothesis of full myopia is rejected, especially since the variable capturing the size of the individual deductible, or current price of care, is very small in magnitude and not statistically significant across most of the estimations. These findings are robust across all approaches, including the results of the sensitivity analyses. However, the overall marginal effects show that while the labor women may spend more in response to the lower EFP, their overall demand is lower than the women in the non-labor group, which is likely driven by other health-associated factors, most notably maternal age. Hence, while giving birth affects the demand for care through price, it also is likely to influence it as a health-related covariate.

One finding that does not support the idea of fully forward-looking behavior is the outcome of Approach III that considers only the labor women and their response to the month of labor. It shows that when labor happens later in the year, spending on non-pregnancy care is lower. This may present evidence of some myopia as the month of labor should not matter for the women in the labor group since their EFP will be lowered regardless of when they give birth. Interestingly, these results are positive and significant for the utilization outcome, potentially capturing the time and other opportunity costs associated with getting medical care around the time of giving birth. Overall, these findings indicate a mixture of both forward-looking and myopic behavior in the demand for medical care, similar to the findings of Aron-Dine et al. (2012). Labor as a medical condition is likely to have a direct effect on the demand for medical care not just through the price effect, but through some health and opportunity costs effects, especially on utilization of medical services.

Finally, the methodological exercises conducted in this part of the study contribute to the growing literature on modeling healthcare costs by examining the data in this very specific sample and applying various estimation techniques. Researchers in this field agree that no one econometric model is more suitable for these data; a careful investigation of the distribution and other aspects of a particular study sample are required before choosing a model. Several estimators were applied to the expenditure and utilization outcome variables in this section. The most notable finding comes in the calculation of marginal effects, where the correct choice of an estimation method is important to make appropriate conclusions.

Chapter 7

EFP in the Demand Equation

7.1 Introduction and Motivation

The estimations and results presented in Chapter 6 examine annual demand for non-pregnancy care among pregnant women where labor costs are expected to lower the expected end-of-year price. Under this approach the EFP is not explicitly incorporated into the demand estimations, but is proxied by the labor variable. This chapter provided interesting insights on the anticipatory effects of nonlinear pricing, but also found that the labor variable in the given context may not be the best proxy or instrument for the price change because the overall impact gets "muddled" by the particular health effects associated with pregnancy and labor. Moreover, about 65% of non-labor women have out-of-pocket spending higher than their deductible, indicating that labor may not be the only health event that moves pregnant women past the deductible threshold into the portion of the budget constraint with lower price.

A few recent studies using data for Medicare Part D enrollees (Jung et al., 2014a and 2014b) examined anticipatory effects in the demand for medical care under nonlinear pricing created by the gap coverage structures by explicitly incorporating the EFP into the estimation of demand for prescription drugs. Another study by Aron-Dine et al. (2012), using the MarketScan data, estimated this price explicitly and used it to examine the demand for medical care in the general sample of employed individuals. Even though this approach is potentially more "noisy" because of estimating the EFP and including this estimated price in the demand equation, the findings of the previous chapter show that this approach may be another way to look at the anticipatory behavior in this particular setting.

This chapter uses the same sample: pregnant women without pregnancy complications. However, labor is not used as the main price-lowering event, but as a health-related covariate. Without explicitly distinguishing between the labor and non-labor groups, the central question examined in the empirical work is whether consumers are myopic or forward looking in their demand for medical care. Two research questions, similar to the ones posed and examined in Chapter 6 are investigated:

- Do women who reach the deductible, either due to giving birth or as a result of other medical spending, have higher demand for medical care?
- Do they use expected end-of-year price when making purchasing decisions?

The answer to both questions is obtained by examining the probability of reaching the deductible and constructing the expected end-of-year price as a function of this probability. The effect of this price on demand would provide evidence on the anticipatory behavior in the presence of deductibles and show whether demand is responsive to reaching the deductible, i.e. to the price nonlinearities. The effect of labor on demand is still of interest here due to the nature of the sample used in the study.

The main assumption here is that all women in the sample begin the calendar year with the similar spot price: their individual deductible. However, their future prices are different, as a result of different demand for care throughout the year. This demand is affected by labor and other factors. The hypothesis guiding the research is that pregnant women who hit their deductible consume more pregnancy and non-pregnancy related care. The central empirical set-up consists of two stages. First, the EFP of each woman is estimated as a function of the probability of reaching the individual deductible. Next, the demand for care is examined as a function of the EFP and other factors.

7.2 Empirical Set-up

The objective of this part of the study is to investigate the demand for medical care in response to the expected end-of-year price, which reveals anticipatory behavior in the presence of nonlinearities. The first step in the analysis is to construct each woman's

EFP. The EFP is calculated as a function of whether a woman reached the deductible (similar to Ellis and McGuire, 1986, Aron-Dine et al., 2012 and Jung et al., 2014a and 2014b). As discussed in Section 6.3, the EFP is likely endogenous in this case because it may be determined simultaneously with the spending (or utilization). Hence, the probability of reaching the deductible must be estimated using some identifying variables:

$$\text{Prob} = f(\mathbf{X}, \mathbf{Year}, \mathbf{S})$$

where Prob is a binary variable (=1 if deductible was reached or exceeded, and =0 otherwise), $\mathbf{X}$ are the demand-shifters that effect the EFP, and $\mathbf{S}$ are the variables that shift the price for enrollees with the same demand for care. These are the identification variables that would deal with the simultaneity problem. The simultaneity problem may arise as high spending before reaching the deductible would lead to low expected prices by increasing the probability of reaching the deductiblein, while the spending before reaching the deductible would in turn depend on these expected prices. Jung et al. (2014a and 2014b) use variable that describe variation in insurance benefits, i.e. insurance characteristics, as shift variables in $\mathbf{S}$ to predict the probability of hitting the deductible and then construct the predicted EFP for each woman. These predicted prices would capture the individuals' expectations about their EFP and also address the simultaneity problem: high-spending women are likely to have low end-of-year prices. The variables included in $\mathbf{X}$ are the same demographic variables as used in the previous chapter. **Year** includes year fixed-effects.

The probability of reaching the deductible is estimated with a logit model:

$$Prob_{ijt} = \frac{1}{1 + \exp(\gamma \mathbf{X}_i + \delta \mathbf{Insur}_{ij} + \tau \mathbf{Year}_t)} \tag{7.1}$$

Based on the construction of the dataset, all women who give birth in the calendar year are expected to reach the deductible, so their $Prob = 1$. The expected end-of-year price (EFP) is then constructed as

$$EFP = (1 - Prob) + Prob \times c$$

where c is the coinsurance rate. This construction follows the definition of p^e in Chapter 3 and Figure 1, where the price after reaching the deductible is the coinsurance rate. The first part of the equation, $(1 - Prob)$ accounts for the first part of the budget constraint in Figure 1 (where $p = 1$), and the second part describes the second part where $p = c$. So, for the women in the labor group, $EFP = (1 - 1) + 1 \times c = c$, which is how this study describes the price for women who give birth in the calendar year and reach their deductible with certainty. The vector of the insurance characteristics used as shift variables $\mathbf{S}$ are family level deductible, coinsurance rate, family out-of-pocket maximum, preventative care coverage, and copayment amounts for primary care, specialist visits, ER, and inpatient care.

The second step in the analysis is the demand estimation:

$$y_{ijt} = \alpha + \beta EFP_{it} + \theta b_t + \lambda bm_t + \gamma \mathbf{X}_i + \tau \mathbf{Year}_t + \varepsilon_{ijt} \tag{7.2}$$

where all the variables are defined as above. "Labor" and "months before labor" are added as health covariates. The main coefficient of interest here is β. The direction, magnitude, and statistical significance of the EFP variable would provide answers to the questions about the demand responsiveness posed above. Additionally, it would show whether the women in the sample take this price into account when making consumption decisions about medical care. The underlying hypothesis is that since reaching the deductible lowers the expected end-of-year price, then forward-looking women would "respond" to this price by consuming more medical care. Empirically, this approach answers the two research questions in "one step" through the coefficient on the EFP variable, compared to the approach used in Chapter 6 where several coefficients were considered to answer the research questions and draw conclusions on anticipatory behavior. The coefficient on the labor variable is also of interest here because of the study sample of pregnant women and findings in Chapter 6 about the impact of labor on the demand for medical care. Thus, its effect is also examined in this chapter because of its central importance to this study.

The empirical approach used here is similar to Section 6.2. The same sample and methods as in Chapter 6 are used. Annual non-pregnancy related expenditures and utilization are examined. Only non-pregnancy related care is used to isolate the more discretionary care among pregnant women, for the same reasons as discussed in Chapter 6.

7.3 Results

The main idea behind this approach is that women who reach their deductible lower their expected end-of-year price and thus spend more on medical care. Using annual expenditures and utilization as the outcomes of interest allows to investigate whether women are myopic or forward-looking: if women are forward-looking (i.e. respond to their expected end-of-year price), their demand for medical care should not depend on the timing of expenses. The sample used for the estimations in this chapter is the same as in Section 6.2: normally pregnant women enrolled in a PPO plan with a deductible. The summary statistics for this sample are presented in Chapter 5 in Table 4.2 and Table 4.3. The outcome variables are annual spending on non-pregnancy related services and annual utilization of these services.

The first step of the analysis is to estimate the probability of reaching the deductible. About 65% of women who did not give birth reached their deductible. Table 7.1 presents the results of this estimation. Age, family size, and the type of employee have a significant effect on this probability. Coinsurance rate has a negative impact, which is intuitive: higher coinsurance rate, which is the price of care after the deductible is met, is a deterrent to reaching the deductible since the price change would be smaller in the case of high coinsurance. Copayment amounts, except for inpatient care, are not a significant determinant of reaching the deductible. Preventative care coverage has a negative significant effect, which is not surprising: higher probability that some care is covered as preventative (and hence not subject to the deductible) leads to lower probability of reaching the deductible.

Using these predicted probabilities, the EFP is constructed. Table 7.2 summarizes these estimates. The predicted probability of reaching the deductible is about 90% on average, and is actually about 81% for the non-labor group. The EFP is 0.22 for the general sample, and, as expected, much lower for the labor group. In fact, it is over twice as large for the non-labor group.

Table 7.3 shows the results of the demand estimations for the expenditure variable. Three empirical methods are used for the estimations, similar to the analyses of the annual demand with labor as a proxy for EFP. Although there are no noteworthy differences between the direction of the OLS coefficients and the GLM/EEE coefficients, based on

Table 7.1: Probability of Reaching the Deductible

	Probability of Reaching Deductible
Age	0.22474***
	(0.02105)
Age squared	-0.00330***
	(0.00035)
Family size	-0.03523***
	(0.00546)
Relationship to employee	0.00433
	(0.02935)
Type of employee	0.02320***
	(0.00522)
Family deductible	-0.00073***
	(0.00003)
Coinsurance rate	-0.88247***
	(0.20980)
Family OOP maximum	-0.00002**
	(0.00001)
Preventative care coverage	-0.22391***
	(0.05605)
Copayment: primary care	-0.00370
	(0.00315)
Copayment: specialists	0.00018
	(0.00205)
Copayment: ER	-0.00001
	(0.00042)
Copayment: inpatient care	0.00079**
	(0.00034)
Constant	-1.10204***
	(0.31891)
Observations	38808

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7.2: Probability of Reaching Deductible and EFP

	All	No Labor	Labor
Probability of reaching deductible	0.90	0.81	1
	(0.12)	(0.11)	(0.00)
EFP	0.22	0.29	0.13
	(0.13)	(0.12)	(0.08)

Standard deviation in parentheses

the discussion in Section 5.2, the GLM and EEE estimates are used to make conclusions and calculate the elasticities. The coefficient on EFP is negative and significant across all three models, indicating that women respond to their expected end-of-year price: higher price leads to lower spending. Labor has a positive significant effect on demand, confirming the findings from the previous chapter. Interacted with age and family size, labor has a negative significant effect, once again pointing to some additional health or time-related effects of labor on demand. Using the GLM and EEE results, women in the labor group spend about \$233 less than those in the non-labor group—this is the overall marginal effect of labor.[1] This finding is consistent with the results from Section 6.2: while labor has a positive price effect on demand, when combined with other factors it actually leads to lower spending on non-pregnancy care. The price elasticity is -0.11 (with the range of -0.06 – -0.16). These are somewhat lower than the elasticities found for the annual spending in Section 6.3, but are consistent with the findings in the literature.

Table 7.4 presents the results of the demand estimations for the utilization variables, using the same methods as in Section 6.2 for the utilization variables: the Poisson regression, the Negative Binomial regression, and a Nonlinear Least Squares method based on the ECM structure. The findings are similar to the results for the expenditures and are similar across the three methods used. The EFP is negative and significant, while the labor variable is positive and significant. Combined with the effect of age and family size, the women in the labor group have on average about 1.4-1.5 fewer visits than the non-labor women. The price elasticity in this case is -0.11– -0.12, ranging from -0.09 to -0.14 (robust across all three methods). This is a similar elasticity to the one found

[1] Note that the OLS marginal effect of labor is much larger in this case: it is about \$748.

Table 7.3: Annual Demand: Expenditures

	OLS	GLM	EEE
EFP	-0.62897***	-0.51201***	-0.51519***
	(0.07454)	(0.12201)	(0.12201)
Labor	0.06676	0.45167***	0.44706***
	(0.08578)	(0.14993)	(0.15161)
Months before labor	-0.01043***	-0.05192***	-0.05239***
	(0.00349)	(0.00637)	(0.00678)
Age	-0.12990***	-0.19923***	-0.19902***
	(0.01206)	(0.02149)	(0.02273)
Age squared	0.00268***	0.00399***	0.00398***
	(0.00020)	(0.00036)	(0.00039)
Labor×age	-0.01668***	-0.02072***	-0.02054***
	(0.00269)	(0.00467)	(0.00481)
Family size	-0.00717*	-0.00294	-0.00293
	(0.00378)	(0.00655)	(0.00653)
Labor×family size	0.00335	-0.00211	-0.00206
	(0.00546)	(0.00920)	(0.00923)
Relationship to employee	0.04715***	0.03780	0.03793
	(0.01488)	(0.02670)	(0.02674)
Type of employee	0.02189***	0.01559***	0.01553***
	(0.00243)	(0.00422)	(0.00422)
Constant	7.98643***	9.73630***	2.65384***
	(0.18632)	(0.33505)	(0.34657)
Observations	38808	38808	38808

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

in Section 6.3 for utilization and similar to the elasticities for the expenditure variable found in this chapter.

Overall, the outcomes of the estimations of annual demand for non-pregnancy related care provide evidence that the women in the sample respond to their expected end-of-year price—the coefficient on the EFP variable is negative and significant across all empirical methods and for both spending and utilization. These results show that women who have a lower probability of reaching the deductible (and hence face a higher EFP) spend and use less non-pregnancy related care in the calendar year. This robust finding points to the presence of forward-looking behavior among the women in the sample. Giving birth has a similar effect on the demand as found before when labor (rather than the probability of reaching the deductible) was used as the price-changing factor: while it has a positive effect on demand (likely through the price lowering influence), when combined with other factors, most notably age, it actually leads to lower spending and utilization.

7.4 Sensitivity Analyses: Plans with Higher Deductibles

The premise for the analysis in Sections 7.3 is that consumers use their expected future price when making decisions to use medical care, and this future price is constructed via the predicted probability of reaching the deductible. It is assumed that normally pregnant women who do not give birth in the calendar year may reach the deductible since they are not expected to have major medical expenditures. However, the results show that their estimated probability of reaching the deductible is quite high: about 81% for the general sample, and 91% for the smaller trimester subsample. A possible explanation for this is that the average size of deductibles is not very high: it is about \$363 for the general sample, and \$345 for the trimester subsample. Compared to the average out-of-pocket expenditures, which are about \$938 on average, and \$498 for the non-labor group, the deductible amount may not be of magnitude that would pose a serious price barrier for consumption of medical care, whether general care or non-pregnancy related. In fact, the average out-of-pocket spending on non-pregnancy related care is \$267, and \$345 for the non-labor group.

To examine the robustness of the findings, it is useful to test the response to higher

Table 7.4: Annual Demand: Utilization

	POI	NBR	ECM
EFP	-0.54283***	-0.54901***	-0.52719***
	(0.05988)	(0.05818)	(0.06471)
Labor	0.15005**	0.12103**	0.18951***
	(0.06213)	(0.06060)	(0.06782)
Months before labor	0.01033***	0.01094***	0.00943***
	(0.00240)	(0.00237)	(0.00252)
Age	-0.04966***	-0.05229***	-0.04509***
	(0.00917)	(0.00892)	(0.01006)
Age squared	0.00128***	0.00131***	0.00121***
	(0.00015)	(0.00015)	(0.00017)
Labor×age	-0.01273***	-0.01178***	-0.01383***
	(0.00200)	(0.00194)	(0.00217)
Family size	-0.00578*	-0.00485	-0.00696*
	(0.00334)	(0.00314)	(0.00372)
Labor×family size	0.00443	0.00434	0.00399
	(0.00431)	(0.00411)	(0.00470)
Relationship to employee	-0.00650	-0.01020	-0.00100
	(0.01152)	(0.01112)	(0.01276)
Type of employee	0.01285***	0.01309***	0.01269***
	(0.00187)	(0.00180)	(0.00201)
Constant	2.43213***	2.47796***	2.35058***
	(0.14217)	(0.13852)	(0.15458)
Observations	38808	38808	38808

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

deductibles that could present a serious price consideration.[2] To isolate the subsample of women who face a higher deductible, only those who have deductibles higher than the average deductible for that calendar year are kept. This new subsample has 16,650 women, which is less than half of the general sample. About 45% of them are in the labor group. The average deductible size is $515.

The two-step procedure described in Section 7.2 is used. First, the probability of reaching the deductible is estimated, followed by the construction of the EFP. The demand equation is then estimated for both the annual expenditures and trimester spending. Only expenditure variables are used in these estimations, with GLM as the main empirical method. Table 7.5 below shows the estimates of the probability of reaching the deductible and the EFP. About 75% of the non-labor women have the probability of reaching the deductible, which is lower than found previously.The EFP is lower for the labor women, as expected, and the difference between the prices faced by the labor and non-labor women is larger than in the general sample.

Table 7.5: Probability of Reaching Deductible and EFP: "High Deductible" Sample

	All	No Labor	Labor
Probability of reaching deductible	0.86	0.75	1
	(0.16)	(0.13)	(0.00)
EFP	0.25	0.35	0.13
	(0.16)	(0.14)	(0.09)

Standard deviation in parentheses

Table 7.6 presents the results of the demand estimations. The first column shows the findings for the annual expenditures. The EFP has a negative significant impact, as in the estimations on the full sample. The coefficient on the labor variable is positive and significant. Table 7.7 presents the price elasticities and the marginal effects of labor. For the annual expenditures, the price elasticity is -0.12, which is similar to what was found in Section 7.3. Giving birth decreases the annual spending on non-pregnancy care by about $292.

[2] The data set includes observations from 1996-2009, a span that saw gradual but consistent increases in the deductible sizes. Since the study uses only those in the PPO plans, the enrollees in the consumer-driven health plans (CDHP), present in the last few years of the dataset, are not included in the study, due to the potential selection problems associated with using different types of plans.

Table 7.6: Demand Estimations: Sensitivity Analyses

	GLM
EFP	-0.48563***
	(0.17066)
Labor	0.51734**
	(0.21976)
Months before labor	-0.05524***
	(0.00887)
Age	-0.15223***
	(0.03084)
Age squared	0.00320***
	(0.00051)
Labor×age	-0.02345***
	(0.00645)
Family size	0.00546
	(0.01027)
Labor×family size	-0.00784
	(0.01382)
Relationship to employee	0.03225
	(0.04132)
Type of employee	0.02964***
	(0.00779)
Constant	8.97246***
	(0.49208)
Observations	16550

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Overall, the results of the sensitivity analyses confirm the findings from above, in direction, significance, and magnitude. Therefore, facing a higher deductible in the insurance policy has a similar effect on the demand for care as a lower deductible. Expected future price matters in the consumption decisions, while labor has an overall negative effect. Hence, the size of the deductible may not matter for consumer decisions, while the expected price does. This echoes the findings in Chapter 6 where the size of the deductible was included in the demand estimations as a proxy for current price. Even though this chapter estimates and explicitly includes the end-of-year price in the demand equations and treats labor as a health covariate, not as a price-changing factor, these similar findings about the size of the deductible corroborate earlier results and provide more insight to reject the hypothesis of full myopia: it appears that the women in the sample use expected future rather than current price when making decisions about medical care consumption.

Table 7.7: Marginal Effects and Elasticities

	GLM
Price elasticity	-0.12***
	(0.04)
Marginal effect of labor	-292.17***
	(83.09)

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

7.5 Discussion

This chapter provides additional evidence on the importance of considering nonlinear price structure created by the deductibles in a health insurance policy and gives new insights on the anticipatory behavior in the demand for medical care. It poses similar research questions as Section 6: (1) Do consumers who reach the deductible have higher demand for medical care?; and (2) Do they use current or expected end-of-year prices when making purchasing decisions? It builds on the empirical strategies and findings of Chapter 6, but takes a different approach to answering these questions. Instead of using labor as a the price-changing factor and not including the expected future price explicitly in the demand equations, this part of the study estimates this price by predicting the probability of reaching the deductible for every woman, regardless of whether she gave birth in the calendar year, and then includes this estimated EFP into the examination of medical care spending and utilization.

The empirical strategy is based on a two-step process: constructing the EFP via the probability of reaching the deductible for the non-labor group (since the labor group reaches it with certainty) and then estimating the demand equation. The same sample of normally pregnant women as used in Chapter 6; their annual demand for non-pregnancy related care in response to their EFP is examined.

The findings reveal that women in the sample indeed take into account their expected end-of-year price in making decisions about consumption of medical care. The coefficient on the EFP variable is negative and significant across most of the estimations, and the price elasticities are in the range of -0.08 – -0.12, which is similar to the findings in Chapter 6 and consistent with other studies that use the price nonlinearity created by the deductibles. In summary, these results show that those who reach the deductible have a higher demand for care as a result of facing a lower expected end-of-year price. The size of deductible, or the current price, may not be a significant factor for the women, as shown in the sensitivity analyses, which also highlights the robustness of the results.

Even though the fact of giving birth is of lesser importance for this chapter than Chapter 6, it is still of interest because of the particular nature of the study sample. The price effect of labor, found in Chapter 6, is incorporated in the construction of the EFP, and labor is included as a health covariate. Labor is once again found to have a positive effect on the annual demand, but when combined with other factors such as age and family size, it actually lowers spending and utilization. These findings also corroborate the results of Chapter 6 and further point to the fact that labor may influence demand through channels other than price.

Finally, the application of different econometric methods in this chapter further highlights the importance of choosing the appropriate methods for estimating specific samples of healthcare data. The outcomes of the estimations done by the least squares methods show the dangers of over-relying on these standard techniques for transformed dependent variables. Generalized linear models, standard and extended, appear to be better suited for the continuous dependent variables in these data. For the count data, traditional methods such as Poisson and Negative Binomial, appear to work well and produce similar results to the regression methods based on the exponential conditional mean structure.

Chapter 8

Deductibles and Trimester Demand

8.1 Introduction and Motivation

Chapters 6 and 7 test the study hypotheses and examine forward-looking behavior using yearly non-pregnancy related expenditures. This annualized spending can occur any time during the year. For some women, the majority of this spending happens during their pregnancy, but for others, only part of this spending happens during pregnancy, if they are only pregnant for a small part of the calendar year. While labor is considered as the main price changing event in Chapter 6, and the probability of reaching the deductible does the same in Chapter 7, the application of both approaches to testing price response for yearly expenditures may be capturing other health states that happened outside pregnancy, and hence may overstate or understate the response to the price change. Narrowing the spending period to only the pregnancy period eliminates other health states and provides a homogenous health condition eliminating various sources of heterogeneity.

In other words, using annual spending may be too broad if the goal is to examine consumption behavior in relation to price change during a homogenous health state. For the purpose of this study, it may be useful to narrow the spending period by isolating non-pregnancy related expenditures that happened while a woman was pregnant. Thus, spending during pregnancy is used to test for forward-looking behavior and examine the demand response to the expected future price. Total spending during pregnancy (i.e. spending during the first, second, and third trimesters) is constructed and used as the outcome variable. In addition, demand for non-pregnancy related care during individual trimesters is also estimated to compare price response during narrower time periods and fairly homogenous health states across labor and non-labor group. Overall, this setting can be viewed as a grounds for quasi-sensitivity analyses for the findings in Chapters 6 and 7.

There is additional rationale for looking at each trimester separately. The idea behind isolating first trimester demand is similar to the premise of the Aron-Dine et al. (2012) study. They test for forward-looking behavior by examining initial utilization of medical care by people who join health plans in different months of the year and hence face different probabilities of reaching the deductible before the end of the calendar year. During the first trimester pregnant women learn about the potential price change associated with labor or other expenditures. This period can be used in the same way as in the Aron-Dine et al. study to examine initial utilization and whether and how women respond to the EFP during that period. If they are forward-looking, they would respond to the EFP in their first trimester if they are anticipating a price change due to reaching the deductible.

However, first trimester is often viewed as somewhat precarious pregnancy period, during which a woman may choose to forgo any additional medical care due to concerns about the proper development of her baby. Hence, expected future price may not play a significant role in her medical care consumption decisions. Moreover, many women do not realize that they are pregnant until the middle of their first trimester. Therefore, they are not aware of the future change in price and may not adjust their demand accordingly during the initial weeks of pregnancy.

Second trimester may be a better period to test anticipatory price effects. In addition to being medically homogenous for most normally pregnant women, regardless of year of giving birth, the women in their second trimester are fully aware of the upcoming future price change due to labor (which may happen during the year of pregnancy or in the next calendar year). The third trimester spending is subject to similar logic as the second trimester: women are aware of the upcoming change in price, and hence should have higher demand for non-pregnancy related care, if they are forward-looking. However, as the third trimester progresses towards the birth, pregnant women typically have more frequent pregnancy-related visits, and hence may be time constrained for non-pregnancy care.

The main research questions remain the same as in Chapters 6 and 7:

- Do women who reach their deductible in a calendar year, either via giving birth or via other expenditures, have higher demand for non-pregnancy related care?
- Which price, CP or EFP, do they use when making medical care consumption decisions?

8.2 Sample Construction

The cohort of women used for this analysis is a subsample of the main sample used in Chapters 6 and 7. Due to the nature of the insurance claims dataset used in this study, each woman may or may not be in the data for consecutive years, so the data are not a complete panel. To isolate the pregnancy period and identify separate trimesters, the month of labor has to be in the data so that the first month of pregnancy can be identified by counting backwards since it cannot be identified from medical claims. There are some women who are in the data for two consecutive years: year of pregnancy and year of birth. It is straightforward to isolate the first month of pregnancy for these women. For those women who are in the data only for one year, it is possible to identify the first trimester only if they got pregnant and gave birth in the same year, which limits this group to only those who give birth in the fourth quarter of a calendar year, i.e. in October, November, and December. Thus, the trimester estimation sample is constructed as follows. The first part of the sample is a group of women who are represented in the sample for two consecutive years, and gave birth in January-September—they constitute a panel with a year of pregnancy and year of birth. The second part of the sample includes women who are in the dataset for one year and give birth in October-December.

First month of pregnancy is defined as (month of labor $-$ 9), and each trimester is defined by adding together consecutive three-month periods. Table 8.1 shows the construction of the trimester periods for each "first month of pregnancy" group, or the pregnancy cohorts. Several pregnancy cohorts, identified by the first month of pregnancy, have trimester periods that overlap years of pregnancy and birth. For example, those who got pregnant in November and December of year t (i.e. year when labor=0), have the first trimester period that includes months in both years t and $t+1$. Thus, their first trimester spending is excluded from the estimations because they face different prices during their first trimester. Second trimester spending is excluded for women who got pregnant in August and September of year t. Third trimester expenditures are not calculated for women who got pregnant in May and June of year t.[1]

Table 8.1: Trimester Sample Construction

First Month of Pregnancy	Labor month	Labor Year	Trimester 1 Months	Trimester 2 Months	Trimester 3 Months
1	10	t	1, 2, 3	4, 5, 6	7, 8, 9
2	11	t	2, 3, 4	5, 6, 7	8, 9,10
3	12	t	3, 4, 5	6, 7, 8	9, 10, 11
4	1	$t+1$	4, 5, 6	7, 8, 9	10, 11, 12
5	2	$t+1$	5, 6, 7	8, 9, 10	
6	3	$t+1$	6, 7, 8	9, 10, 11	
7	4	$t+1$	7, 8, 9	10, 11, 12	1, 2, 3
8	5	$t+1$	8, 9, 10		2, 3, 4
9	6	$t+1$	9, 10, 11		3, 4, 5
10	7	$t+1$	10, 11, 12	1, 2, 3	4, 5, 6
11	8	$t+1$		2, 3, 4	5, 6, 7
12	9	$t+1$		3, 4, 5	6, 7, 8

Table 8.2 shows how the labor and non-labor groups are defined in the trimester estimations. Each pregnancy cohort has trimester periods in either year t or year $t+1$, so the number of pregnancy cohorts in the labor and non-labor groups changes with different trimesters. For example, the first trimester has 7 non-labor cohorts and 3 labor cohorts. The second trimester has 6 labor cohorts, and 4 non-labor cohorts. The third trimester has only one cohort in the non-labor group, and 9 cohorts in the labor groups.

[1] Since the normal gestation period is 10 calendar months, it is used here to identify the month of conception. It is an approximation to the true date of conception, so the trimesters calculated based on the estimated first month of pregnancy may or may not correspond directly to the actual course of pregnancy experienced by each woman. Given the limitations of the claims data, this is the best approximation to the true trimester periods.

Thus, third trimester expenditures are not estimated due to the differences in the size of labor and non-labor groups.

To construct trimester-level expenditures, monthly expenditures are aggregated based on the trimester construction in Table 8.1. Missing data on trimester expenditures are recoded to zero only if a woman is identified to be pregnant during that trimester period, to reflect her decision not to consume medical care during the period with missing data. Thus, two-part models of the type described in Section 5.3.4 can be used for the trimester estimations since the zero expenditures reflect the decision not to purchase care, and are not due to missing values. For this reason, the number of observations for each trimester is different. The expenditures for each trimester are aggregated on an annual basis to construct the yearly non-pregnancy related spending that occurred during pregnancy. For example, if a woman got pregnant in June of year t, her annual trimester expenditures for year t would be the sum of her expenditures during the first and second trimesters. She would not have any trimester expenditures for year $t+1$ since her third trimester overlaps both year t and $t+1$ as discussed above. On the other hand, the yearly trimester expenditures for a woman who got pregnant in October of year t would equal to first trimester spending in year t and the sum of expenditures during second and third trimesters in year $t+1$.

Table 8.2: Labor and Non-Labor Groups by Pregnancy Cohort

Trimester	Labor Group (Labor=1)	Non-labor Group (Labor=0)
Trimester 1	1, 2, 3	4, 5, 6, 7, 8, 9, 10
Trimester 2	1, 2, 3, 10, 11, 12	4, 5, 6, 7
Trimester 3	1, 2, 3, 7, 8, 9, 10, 11, 12	4

This subsample has 9,253 observations, with about 84% of women giving birth in the calendar year. Table 8.3 shows the descriptive statistics for this sample. The demographic characteristics are similar to those of the main sample. The average size of the deductible is \$368 and is slightly lower for the labor group compared to the non-labor women. The average trimester spending on non-pregnancy related care is about \$357 for this sample. These expenditures are higher for the labor group.

Table 8.3: Descriptive Statistics: Trimester Sample

	All Sample	No-labor	Labor
Labor	0.84	-	1.00
	(0.37)	-	(0.00)
Month of labor	9.41	-	9.41
	(2.02)	-	(2.02)
Age	29.23	28.99	29.28
	(5.12)	(4.97)	(5.15)
Family size	4.21	4.37	4.18
	(2.56)	(2.63)	(2.54)
Relationship to employee	0.49	0.44	0.50
	(0.50)	(0.50)	(0.50)
Individual deductible ($)	368.47	397.84	362.73
	(223.41)	(288.71)	(207.81)
Family deductible ($)	853.65	891.28	846.29
	(521.35)	(607.31)	(502.55)
Coinsurance (%)	0.14	0.15	0.14
	(0.08)	(0.07)	(0.08)
Individual OOP maximum ($)	1912.98	1989.01	1898.12
	(1108.60)	(1061.79)	(1116.99)
Family OOP maximum ($)	3660.08	3819.60	3628.90
	(2174.57)	(1818.03)	(2236.41)
Preventative care coverage	0.85	0.99	0.82
	(0.36)	(0.08)	(0.38)
Copayment: primary care ($)	4.70	3.26	4.97
	(8.24)	(7.03)	(8.43)
Copayment: specialists ($)	5.65	3.96	5.97
	(11.02)	(9.22)	(11.30)
Copayment: ER ($)	40.36	38.05	40.80
	(38.95)	(33.63)	(39.86)
Copayment: inpatient care ($)	6.25	8.21	5.88
	(37.91)	(41.76)	(37.13)
First trimester expenditures ($)	223.61	223.90	223.50
	(564.13)	(506.20)	(584.87)
Second trimester expenditures ($)	132.54	116.05	134.59
	(552.38)	(308.00)	(575.57)
Third trimester expenditures ($)	131.25	121.38	131.45
	(462.15)	(193.55)	(466.13)
Total trimester expenditures ($)	357.12	303.93	367.52
	(875.58)	(587.73)	(921.07)
Number of trimesters	2.15	1.68	2.24
	(0.85)	(0.66)	(0.85)
Observations	9253	1513	7740

Note: Standard deviations in parentheses.

8.3 Empirical Set-up

Both empirical approaches, from Chapter 6 and Chapter 7, are used to examine the trimester demand in the presence of deductibles. The first one uses labor as the price changing health event. Table 8.1 shows how the labor and non-labor groups are constructed. The main premise behind it is comparing demand for non-pregnancy related care between two groups: pregnant women who give birth in the calendar year and thus know that they will face a lower EFP by exceeding the deductible, and those who give birth in the following calendar year and thus face a higher end-of-year price. Controlling for the timing when these expenditures occur, i.e. only using spending during pregnancy, holds constant medical factors that might have biased comparisons done on the overall annual expenditures. The demand equation is

$$trimexp_{ijt} = \alpha + \beta b_t + \lambda tc_t + \gamma \mathbf{X}_i + \delta \mathbf{Insur}_j + \theta \mathbf{Insur}_j \times b_t + \tau \mathbf{Year}_t + \varepsilon_{ijt} \quad (8.1)$$

where *trimexp* is the expenditures during all trimesters in year t. For the estimations for separate trimesters, *trimexp* becomes the variable for spending in that particular trimester ($trimexp_1$, $trimexp_2$, $trimexp_3$). A variable "count of trimesters" (tc) is added to capture the number of trimester periods that a woman has in the year that her expenditures are used in the estimations. This variable controls for the length of pregnancy in a particular year.

The second empirical approach is based on estimating the EFP via the probability of reaching the deductible and including the EFP in the demand equation. The EFP is constructed via estimating the probability of reaching the deductible, and the identifying variables are the same as described in Section 7.2. The effect of labor as a price factor is captured in the EFP via the certainty of reaching the deductible. Moreover, in this subsample pregnancy and labor happen either in the same year (with certainty of reaching the deductible) or in two consecutive years (so the full deductible is faced each year).

As common in much of healthcare data, about 50% of women in the sample have zero expenditures and utilization during one of the trimesters, which is expected when analyzing demand for medical care in a shorter interval. There are various suggestions in the health econometrics literature about the fraction of zeros that is considered significant and calls for a two-part model. As described in Section 5.3.4, these models are used when the zeros and positive values arise from two data-generating processes. The zeros come from the initial decision not to use any medical care, and the positive values arise from the second decision about how much care to use by those who chose to use it. The appropriate econometric methods for two-part models are described in Section 5.3.4. Since this study's aim is to examine anticipatory behavior, both decisions—to consume or not, and how much to consume—are of special interest and are examined separately and as part of a two-part model. Logit is used to estimate the first decision: the probability of positive expenditures or use. The second part is based on the positive part of the demand, and an appropriate econometric method is used. For the overall trimester spending, or spending during pregnancy, the number of women with zero expenditures on non-pregnancy related care is less than 10%. Hence, same methods as in Chapter 6, OLS, GLM, and EEE, are used to estimate demand and calculate price elasticicities. The EFP is calculated in the same as in Section 7.2, and the demand equation is:

$$trimexp_{ijt} = \alpha + \beta EFP_{it} + \theta b_t + \lambda tc_t + \gamma \mathbf{X}_i + \tau \mathbf{Year}_t + \varepsilon_{ijt} \tag{8.2}$$

8.4 Results: Labor as Proxy for EFP

Table 8.4 presents the results for the overall expenditures during pregnancy. The labor variable is positive for the log-OLS and GLM models, and negative for the EEE model. This finding reveals that labor, and associated price change, may not play a major role in determining demand for non-pregnancy care during pregnancy. The size of the individual deductible does not have a significant impact on spending. The marginal effects, presented in Table 8.5, are the overall effects that include the effect of the interaction terms between variables. These results show that on average women in the labor group spend about $\$50 - 60$ less than those in the non-labor group. This effect is likely driven by non-price factors, most likely related to health or opportunity costs, as discussed in Chapter 6.

Turning to the spending in individual trimesters, Table 8.6 shows results for the demand for non-pregnancy care in the first trimester. Since almost 50% of these expenditures are zero, the estimations are done in two parts. A two-part model using the -tpm- command in Stata 13 developed by Belotti and Deb (2012) was estimated. The version used here combines logit for the first part and GLM for the second part. The

Table 8.4: Trimester Expenditures

	OLS	GLM	EEE
Labor	0.02012	0.03243	-0.03430
	(0.40796)	(0.46166)	(0.43323)
Trimester count	0.65042***	0.43072***	0.46342***
	(0.02253)	(0.02910)	(0.03292)
Age	-0.02786	-0.16458***	-0.14952***
	(0.03768)	(0.05693)	(0.05266)
Age squared	0.00074	0.00318***	0.00286***
	(0.00062)	(0.00092)	(0.00084)
Labor×age	-0.01064	-0.00619	-0.00319
	(0.01241)	(0.01328)	(0.01203)
Family size	0.02633	-0.00026	0.00148
	(0.02220)	(0.01860)	(0.01869)
Labor×family size	0.02183	-0.00934	-0.01214
	(0.02352)	(0.02133)	(0.02135)
Relationship to employee	0.01104	0.03240	0.31975**
	(0.04474)	(0.04325)	(0.13184)
Type of employee	0.03249***	0.02414***	0.01955**
	(0.00689)	(0.00842)	(0.00862)
Individual deductible	0.00061	0.00001	0.00016
	(0.00081)	(0.00094)	(0.00095)
Family deductible	-0.00023	0.00008	-0.00001
	(0.00040)	(0.00048)	(0.00048)
Coinsurance rate	-2.68234***	-1.80605**	-1.63404**
	(0.89922)	(0.89983)	(0.83356)
Preventative care	0.08283	-0.04800	-0.06969
	(0.07660)	(0.10948)	(0.11827)
Individual OOP max	0.00002	0.00005	0.00002
	(0.00011)	(0.00009)	(0.00010)
Family OOP max	-0.00001	-0.00004	-0.00002
	(0.00007)	(0.00006)	(0.00006)
Constant	3.72259***	7.28343***	1.14676
	(0.64646)	(0.95982)	(0.91396)
Observations	9253	9253	9253

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8.5: Marginal Effects: Trimester Spending

	OLS	GLM	EEE
Labor	20.41	-52.24**	-61.29***
	(24.95)	(23.72)	(24.88)
Individual deductible	0.30	0.14	0.14
	(0.13)	(0.13)	(0.12)

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

results are not presented here because they are a combination of the first and second column of Table 8.6. This model was used to conduct tests of joint significance of the EFP and labor variables. These tests show that the labor and individual deductible variables are not significant across both parts.

The first column has the findings for the logit on the probability of positive spending, i.e. the first part of the model capturing the decision on any or no spending. Labor has a negative effect on the decision to consume any non-pregnancy related care, but it is not significant. The next two columns show the results from the GLM and the ECM approaches on the positive part of the first trimester expenditures. The direction of the coefficients is different, but neither is statistically significant. The size of the individual deductible is not significant and very small in magnitude across all models.

Marginal effects are shown in Table 8.7. Giving birth in a calendar year lowers the probability of having any non-pregnancy related spending by about 5%. This is the only significant result for these estimations. Interestingly, both the GLM and ECM models show a positive dollar effect of labor on first trimester spending (about $17 – $25), but the findings are not statistically significant. For both parts, the marginal effect of labor is about $-\$0.28$ and is not significant. The size of the deductible has a minimal non-significant impact as well.

The second trimester findings are in Table 8.8. Similarly to the first trimester spending, a two-part model is used here to capture both the decision to spend at all and the decision on how much to spend given that the first decision is to buy medical care. The test of joint significance shows that labor is jointly significant at the 5% level, while individual deductible is jointly insignificant across the two parts of the model. The logit

Table 8.6: First Trimester Expenditures

	Logit	GLM	ECM
Labor	-0.12867	0.20362	-0.44338
	(0.47100)	(0.51583)	(0.56161)
Age	0.11815**	-0.10706	-0.18405
	(0.05484)	(0.06629)	(0.12354)
Age squared	-0.00180*	0.00212*	0.00317
	(0.00092)	(0.00109)	(0.00205)
Labor×age	-0.00440	-0.00175	0.02030
	(0.01391)	(0.01488)	(0.01534)
Family size	-0.03260	0.01403	0.01710
	(0.02216)	(0.02395)	(0.02034)
Labor×family size	0.01180	-0.01674	-0.03108
	(0.02596)	(0.02802)	(0.02754)
Relationship to employee	-0.03590	0.08655	0.15603
	(0.06495)	(0.06053)	(0.10170)
Type of employee	0.02254**	0.01518	0.00513
	(0.01147)	(0.00974)	(0.01406)
Ind. deductible	0.00005	0.00059	0.00075
	(0.00084)	(0.00097)	(0.00103)
Family deductible	0.00001	-0.00020	-0.00029
	(0.00042)	(0.00049)	(0.00054)
Coinsurance rate	-1.23954	-1.43162	-1.17147
	(1.03458)	(0.90701)	(1.01961)
Preventative care	0.24468*	0.10757	0.12655
	(0.12937)	(0.11980)	(0.14914)
Individual OOP max	-0.00012	0.00011	0.00009
	(0.00010)	(0.00009)	(0.00009)
Family OOP max	0.00004	-0.00006	-0.00003
	(0.00007)	(0.00006)	(0.00007)
Constant	-0.81350	6.92511***	8.12674***
	(0.87764)	(1.08423)	(1.79015)
Observations	5461	3805	3805

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8.7: Marginal Effects: First Trimester

	Logit	GLM	ECM
Labor	-0.05***	17.14	25.46
	(0.02)	(23.94)	(23.09)
Individual deductible	0.0003	0.19	0.23
	(0.0004)	(0.14)	(0.17)

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

model shows a negative, but not significant impact of labor. However, both the GLM and ECM estimations for the positive significant part of expenditures show a positive impact of labor on spending. None of the models show a significant coefficient for the individual deductible variable.

The marginal effects for these demand estimations are in Table 8.9. Similarly to the first trimester effects, giving birth in a calendar year lowers the probability of any non-pregnancy spending by about 7%. However, women who decide to purchase this type of care during the second trimester would spend about $\$34 - 35$ more if they are in the labor group compared to the non-labor group. Based on the two-part model (i.e. coefficients in the first two columns of Table 8.8) women in the labor group spend about \$9.1 more on non-pregnancy related care during pregnancy than those in the non-labor group, but the result is not statistically significant.

The results for the third trimester estimations are in Table 8.10 where the same two-part model approach is used. It is useful to keep in mind that for these estimations there is only one pregnancy cohort in the non-labor group. The first column shows that labor has a negative effect on the probability of having any positive non-pregnancy related expenditures during the third trimester, which is consistent with the findings for the first and second trimesters. The coefficients on "labor" are positive, but not significant for GLM and ECM models for the positive parts of third trimester spending.[2] The size of individual deductible is not statistically significant.

Marginal effects for the third trimester spending are in Table 8.11. Similar to the

[2] Note that the coefficients for the GLM and ECM models are identical, but the marginal effects are slightly different.

Table 8.8: Second Trimester Expenditures

	Logit	GLM	ECM
Labor	-0.54660	0.92354*	0.92354*
	(0.52086)	(0.47928)	(0.47923)
Age	-0.06747	-0.20291***	-0.20291***
	(0.04360)	(0.06837)	(0.06837)
Age squared	0.00108	0.00399***	0.00399***
	(0.00071)	(0.00117)	(0.00117)
Labor×age	0.00889	-0.02254*	-0.02254*
	(0.01508)	(0.01333)	(0.01333)
Family size	-0.02748	0.00891	0.00891
	(0.02714)	(0.02170)	(0.02170)
Labor×family size	0.02568	-0.02308	-0.02308
	(0.02876)	(0.02705)	(0.02705)
Relationship to employee	0.01123	0.01947	0.00974
	(0.00855)	(0.01514)	(0.06242)
Type of employee	0.01123	0.01947	0.01947
	(0.00855)	(0.01514)	(0.01513)
Ind. deductible	-0.00006	-0.00058	-0.00058
	(0.00102)	(0.00086)	(0.00086)
Family deductible	0.00005	0.00009	0.00009
	(0.00050)	(0.00041)	(0.00041)
Coinsurance rate	-1.30680	-0.75505	-0.75505
	(1.12962)	(1.36234)	(1.36219)
Preventative care	-0.08516	-0.17013	-0.17013
	(0.09598)	(0.20172)	(0.20170)
Individual OOP max	0.00015	-0.00023**	-0.00023**
	(0.00014)	(0.00011)	(0.00011)
Family OOP max	-0.00002	0.00015*	0.00015*
	(0.00009)	(0.00009)	(0.00009)
Constant	1.57412**	7.99970***	7.99970***
	(0.77087)	(1.09673)	(1.09661)
Observations	7894	4485	4485

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8.9: Marginal Effects: Second Trimester

	Logit	GLM	ECM
Labor	-0.07***	35.1*	34.18**
	(0.03)	(21.90)	(15.36)
Individual deductible	0.0002	0.07	0.04
	(0.0002)	(0.14)	(0.11)

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

first and second trimester findings, giving birth in a calendar year significantly lowers the probability of having any non-pregnancy related expenditures by about 3%. For the positive part of expenditures, labor increases the spending by about $\$14 - 15$, but the result is not significant. The marginal effects for individual deductible are very small and not statistically significant. The combined marginal effect of labor for both the first and second part is about \$2, but the result is not significant.

Overall, the results of these estimations based on the approach "labor as proxy for EFP" confirm some of the results in Section 6.2. Giving birth in a calendar year leads to lower demand for non-pregnancy related care during pregnancy. This finding is similar to the results in Section 6.2, but the underlying reasons may be slightly different. Since the labor variable is not significant, and neither are the interaction terms associated with it, the demand response captured here may not be due to price, but other factors. The statistically significant variables in these estimations (Table 8.4) are the trimester count, age, type of employee, and coinsurance rate. The relationship between the trimester count and the expenditures during pregnancy is clear: the more trimesters a woman has in a year, the higher her expenditures are because of having a longer period of time to spend. As discussed above, higher maternal age significantly lowers non-pregnancy (i.e. discretionary) spending because of perceived risks. Higher coinsurance rate, which is applied after the deductible is reached and becomes the price afterwards, has an expected negative effect on demand. This last finding signals to the fact that many of the pregnant women in the sample may reach the deductible regardless of labor, which is why the labor variable does not have a significant impact on spending on its own. However, combined with health-related factors, such as age, and price-related factors,

Table 8.10: Third Trimester Expenditures

	Logit	GLM	ECM
Labor	-1.79529	0.92654	0.92654
	(2.39844)	(0.81001)	(0.80995)
Age	0.13740	-0.20729**	-0.20729**
	(0.08547)	(0.08267)	(0.08266)
Age squared	-0.00281***	0.00412***	0.00412***
	(0.00096)	(0.00137)	(0.00137)
Labor×age	0.03252	-0.01406	-0.01406
	(0.06428)	(0.02499)	(0.02499)
Family size	0.01392	0.03281	0.03281
	(0.14167)	(0.03562)	(0.03561)
Labor×family size	-0.00948	-0.05280	-0.05280
	(0.14222)	(0.03732)	(0.03731)
Relationship to employee	0.02379	-0.04144	-0.04144
	(0.07245)	(0.06416)	(0.06415)
Type of employee	0.01512	0.01285	0.01285
	(0.01211)	(0.01015)	(0.01015)
Individual deductible	0.00333	-0.00071	-0.00071
	(0.00594)	(0.00162)	(0.00162)
Family deductible	0.00075	0.00027	0.00027
	(0.00214)	(0.00074)	(0.00074)
Coinsurance rate	-5.74787	-1.97088	-1.97088
	(7.28370)	(1.73178)	(1.73165)
Preventative care coverage	-0.12099	-0.15136	-0.15136
	(0.13424)	(0.12998)	(0.12997)
Individual OOP maximum	-0.00040	-0.00057***	-0.00057***
	(0.00064)	(0.00017)	(0.00017)
Family OOP maximum	-0.00005	0.00039***	0.00039***
	(0.00035)	(0.00011)	(0.00011)
Constant	1.26903	7.42562***	7.42562***
	(2.53927)	(1.38758)	(1.38748)
Observations	7902	6791	6791

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8.11: Marginal Effects: Third Trimester

	Logit	GLM	ECM
Labor	-0.03***	13.85	15.03
	(0.003)	(20.63)	(15.99)
Individual deductible	0.002	0.05	0.02
	(0.001)	(0.05)	(0.05)

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

such as coinsurance rate, it may lower the demand for discretionary medical care during pregnancy in the year of labor.

The findings on the demand during the individual trimesters offer some additional insights on the relationship between labor and spending. One consistent finding for all three trimesters is that giving birth in a calendar year reduces the probability of having any non-pregnancy related expenditures during any of the trimesters; however, the magnitude of this reduced probability is quite small (between $3-7\%$). What may be more indicative of the relationship between labor and demand is the impact of labor on the positive part of expenditures. The marginal effects of labor for all three trimesters is positive and ranges between \$14 for the third trimester to \$35 in the second trimester. Only second trimester results are statistically significant. These findings indicate that labor, in its price-lowering capacity, leads to higher demand for non-pregnancy related care during individual trimesters. This effect is more pronounced during the second trimester, which is not surprising. Women in their second trimester are both aware of their upcoming price change (vs. those in their first trimester) and are in the "safest" health state during pregnancy to receive any discretionary care (vs. first and third trimesters). This provides additional evidence in favor of forward-looking behavior. Since the impact of the deductible size is not significant (and very small in magnitude) across all estimation, it is clear that the women are not responding to their current price when making consumption decisions.

8.5 Results: EFP in Demand Equation

The second empirical approach, based on Chapter 7, estimates the probability of reaching the deductible, independent of year of labor, constructs the EFP, and then estimates the demand equation with the EFP as the variable of interest. For the trimester sample, only the overall spending on non-pregnancy care during pregnancy is used, to investigate if the women in the sample respond to their expected future price.

Table 8.12 shows the determinants of the probability of reaching the deductible. Out of the demographic variables, age (and its square), family size, and relationship to employee are statistically significant. Coinsurance rate has a negative significant impact on the probability of reaching the deductible, while the effect of copayment amounts for different types of services is not significant. Table 8.13 shows the average probability of reaching the deductible for this sample and the average EFP. This probability is actually quite high for the women who did not give birth, which confirms the deduction based on the findings from the section above that many of the non-labor women may actually reach their deductible. As expected, the average EFP is lower for the labor group.

The results of the estimations for the determinants of the demand for non-pregnancy related care during pregnancy are presented in Table 8.14. The coefficient on the EFP variable is negative across all models, indicating that the women in the sample respond to their expected future price when making consumption decisions. In fact, the price elasticities, presented in Table 8.15, are around -0.09 - -0.10 (statistically significant), based on the GLM and EEE models, which is similar to the findings in Section 7.3. The labor variable is positive and significant in the OLS model, but not statistically significant in the GLM and EEE approaches, which are considered more "trustworthy" based on the nature of these data (see Chapter 5). However, similar to the findings in Section 6.2, giving birth decreases the non-pregnancy related spending by about $\$69-74$. Since labor is included in these estimations as a health-related covariate, it is not expected to have a price effect on the demand.

Overall, the findings on the demand during pregnancy for this empirical approach show that the women in the sample may take into account the expected end-of-year price when making purchasing decisions. The EFP is negative and significant , emphasizing that higher expected end-of-year price deters women from using non-pregnancy

Table 8.12: Probability of Reaching Deductible: Trimester Sample

	Probability of Reaching Deductible
Age	-0.20828**
	(0.09536)
Age squared	0.00378**
	(0.00162)
Family size	-0.05014***
	(0.01921)
Relationship to employee	0.41823***
	(0.10801)
Type of employee	-0.01346
	(0.01891)
Family deductible	-0.00086***
	(0.00009)
Coinsurance rate	-1.51376*
	(0.87678)
Family OOP maximum	-0.00006*
	(0.00003)
Copayment: primary care	0.00487
	(0.01576)
Copayment: specialists	0.00717
	(0.01089)
Copayment: ER	0.00173
	(0.00168)
Copayment: inpatient care	0.00145
	(0.00143)
Constant	11.97489***
	(1.60035)
Observations	8868

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8.13: Probability of Reaching Deductible and EFP: Trimester Sample

	All	No Labor	Labor
Probability of reaching deductible	0.99	0.92	1.00
	(0.05)	(0.08)	0
EFP	0.15	0.22	0.14
	(0.09)	(0.10)	(0.08)

Standard deviation in parentheses

Table 8.14: Expenditures during Pregnancy

	OLS	GLM	EEE
EFP	-1.22608***	-0.72317**	-0.60430*
	(0.25042)	(0.30463)	(0.31463)
Labor	0.79064**	0.07235	-0.01200
	(0.36310)	(0.40852)	(0.34653)
Trimester count	0.65304***	0.42961***	0.46716***
	(0.02248)	(0.02700)	(0.03204)
Age	-0.05114	-0.16843***	-0.14409***
	(0.03858)	(0.05153)	(0.05170)
Age squared	0.00124*	0.00325***	0.00277***
	(0.00064)	(0.00083)	(0.00083)
Labor×age	-0.01900	-0.00619	-0.00328
	(0.01242)	(0.01336)	(0.01128)
Family size	-0.01207	0.00773	0.00580
	(0.02265)	(0.02079)	(0.01883)
Labor×family	0.02074	-0.01909	-0.01857
	(0.02397)	(0.02315)	(0.02161)
Relationship to employee	0.04795	0.00632	0.03206
	(0.04125)	(0.04372)	(0.04276)
Type of employee	0.03790***	0.02705***	0.02155***
	(0.00648)	(0.00700)	(0.00789)
Observations	9253	9253	9253

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8.15: Marginal Effects and Elasticities

	OLS	GLM	EEE
Price elasticity	-0.10***	-0.11***	-0.09 *
	(0.04)	(0.05)	(0.05)
Marginal effect of labor	3.94	-68.95***	-73.74***
	(25.11)	(22.64)	(24.38)

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

related medical care during their pregnancy. The price elasticities are similar to the ones obtained for annual spending that is estimated in Section 7.3. Labor has a negative overall effect on demand for medical care. This is somewhat surprising because normal pregnancy is a fairly homogenous health state for both labor and non-labor women. The potential price effect is incorporated in the EFP based on its construction, which means that labor may have an influence on demand other than through price. Similar to the findings in Section 6.2, the use of medical care in the year of labor may be affected by other factors such as time availability for doctor visits. Overall, the magnitude of this effect is similar to the finding for the annual demand, which shows that the results are robust across various specifications of the outcome variables.

8.6 Discussion

This chapter can be viewed as a quasi-sensitivity analysis to the findings in Chapters 6 and 7. It applies the same approaches as the other two empirical chapters: the innovative approach of using labor as a proxy for the EFP, and the established approach of explicitly incorporating the EFP into the demand equation by calculating it via the probability of reaching the deductible. The first approach is applied both to all non-pregnancy spending during pregnancy (i.e. all trimesters available in a calendar year) and to expenditures in each trimester. The findings echo the results in Section 6.2, but they are not as strong. There is clear evidence against full myopia: the effect of the individual deductible is negligible, both in magnitude and statistical significance. Hence, women do not respond to their current price. The findings of the effect of giving birth in a calendar year are less clear: there is evidence that labor has a positive impact on spending as a stand-alone price variable, but it is not significant across all estimations. The findings are strongest for the second trimester, where labor is seen as increasing demand for non-pregnancy spending. Arguably, this period during a pregnancy is the most suitable testing grounds for the presence of anticipatory effects because women are both aware of pregnancy and hence upcoming price change and are in the safest health state during pregnancy. The results of the second approach, using the EFP in the demand equation, provide clear and strong evidence in favor of forward-looking behavior. The price-elasticities obtained from the estimations are similar to the ones found in Chapter 7, corroborating those results.

Chapter 9

Conclusions and Implications for Policy and Research

This study provides additional empirical evidence on the importance of considering nonlinear pricing when examining demand for medical care. Only a handful of studies, some in the 1980s and others in the recent years, have looked at the demand for care and estimated elasticities based on the dynamic nonlinear, versus static linear, price of care that is created when an insurance policy has deductibles, maximum out-of-pocket amounts, or coverage limits. Moreover, this study makes a contribution to the literature on the anticipatory effects in consumer behavior by examining whether consumers use current or expected end-of-year prices when making purchasing decisions. The nonlinearity in price enables this investigation and allows to make conclusions on forward-looking versus myopic response.

This study uses a sample of normally pregnant women from a comprehensive insurance claims dataset that includes data on medical visits and spending, insurance characteristics, and some demographics, that allows to identify medical conditions, track expenditures and utilization, and link them with the insurance policies of the enrollees. Several checks for selection were conducted to ensure that the sample is constructed very conservatively with little possibility for adverse selection among the women in the sample. Two different approaches were taken to examine demand for non-pregnancy related care and investigate which price, current or expected end-of-year, these women use to make decisions about medical care. Both ask two main questions: (1) Does lower expected end-of-year price lead to higher demand?, and (2) Which price, current or expected end-of-year, do the women use?

The first approach is an innovation of this study. Given the possibilities offered by a claims dataset, a large health event that moves consumers past the deductible threshold can be identified. For the study sample, labor and delivery is such event—it is not only large enough to ensure that the deductible is exceeded and hence expected future price is lowered, it is also predictable and thus provides ideal grounds for the study of anticipatory effects. Moreover, labor is an exogenous health event since the broad timing of it cannot be chosen by a pregnant woman. Some of the women in the sample give birth in the insurance benefit period (a calendar year) and thus face a lower expected end-of-year price, while others do not give birth and may or may not exceed their deductible. Several empirical strategies are employed with labor being the main factor that changes price and annual demand as the outcome variable. The first and simplest uses labor as a proxy for the expected future price. The second uses labor as an instrument for that price. The third uses only the women in the labor group and uses their month of labor to track response to the timing of the price change.

The findings show that giving birth in a calendar year increases spending on and utilization of non-pregnancy related care, confirming that labor affects demand through its price-lowering effect. However, when combined with other factors such as the age of a woman, which is a notable health aspect in pregnancy and labor, labor actually lowers the demand for care. The size of the deductible, which is included as a proxy for the current price of care, does not have a large effect on demand. These results provide strong evidence for the presence of forward-looking behavior. Nevertheless, the null hypothesis of full myopia cannot be rejected. The month of labor estimations show that the women in the labor group may not respond to their end-of-year prices: giving birth later in the year results in lower demand for care, indicating that the women may not respond to their lower end-of-year price. The price elasticity of demand calculated under the "labor as price effect" approach is about -0.21 for expenditure variables and -0.10 for the utilization variables, both of which are within the range of elasticities found in the previous studies that use nonlinear price. Overall, while the findings on the anticipatory effects are slightly mixed, which is similar to the findings of Aron-Dine et al. (2012), there is significant evidence in favor of forward-looking behavior.

The second approach is similar to the those used in a few previous studies that look at nonlinearities created by insurance deductibles and coverage gap in Medicare Part D. Since a high number of women who do not give birth exceed their deductible, it is useful to look at the probability of reaching the deductible as the price-changing "event." Under this approach, the expected future price is calculated using this predicted probability and is included in the demand equation, unlike the first approach where labor (or month of labor) was used as the price changing factor without explicitly incorporating the EFP in the demand estimations. Annual non-pregnancy related expenditures are used for these estimations. The estimation results show that those with lower EFP have higher annual demand for care, and clearly use this EFP when making consumption decisions. The price elasticities are in the range of -0.08 – -0.12, which is similar to what was found under the first approach. Overall, this approach provides additional evidence on the presence of forward-looking behavior: reaching the deductible, and hence facing a lower end-of-year price, leads to higher demand.

Both approaches are applied to a smaller subset of the main estimation sample: a group of women for whom spending in each trimester (and overall spending during pregnancy) can be identified. These estimations are used to isolate the effect of non-linear pricing on a more specific spending interval (pregnancy period rather than the entire calendar year) and provide additional evidence by corroborating the findings of the previous empirical chapters. Using "labor as a proxy for the EFP" strategy, expenditures during the entire pregnancy as well as individual trimesters are examined. The strongest price-lowering effect of labor is seen in its impact on the demand during second trimester—giving birth leads to higher non-pregnancy related medical care spending. The second approach of using the EFP in the demand equation also confirms the previous results—women in the labor group have a lower EFP, and higher expected future price leads to lower non-pregnancy related expenditures during pregnancy. Price elasticities calculated for this sample are in the range of -0.09 – -0.11, which is similar to the elasticities found in other chapters. Overall, these findings further confirm the presence of forward-looking behavior in the sample of pregnant women and reject full myopia.

The empirical exercises conducted in this study and the findings provide valuable insights on consumer behavior and underscore the importance of considering nonlinear pricing in the demand for medical care. The results obtained here show that consumers indeed understand something about the non-linear price they face and respond to their end-of-year prices, which are likely different from current prices for those who have cost-sharing elements in their policies. The price elasticities obtained here are mostly lower than those found in studies that used a single price such as in the RAND Health Insurance Experiment, providing further reasons for caution against using elasticities calculated in this way when insurance policies actually have nonlinear price structures, which echoes the warnings in the literature (see, for example, Aron-Dine et al., 2013).

Furthermore, these findings contribute to the field of behavioral economics by looking at the anticipatory effects in the consumption of medical care. Only one other study, by Aron-Dine and colleagues (2012), previously examined if consumers are myopic or forward-looking when it comes to purchasing medical care. This study is the second to explore these effects by using the expectation of a lower price created by a large predictable health event, which is a unique contribution since other studies in healthcare that considered these effects were mostly on the demand for cigarettes and the price change associated with increase in taxes (see, for example, Gruber and Köszegi (2001), on rational addiction). This study shows that consumers respond to their expected future price and use it to make purchasing decisions, and are thus forward-looking to a nontrivial extent. However, similarly to Aron-Dine et al. (2012) and the studies on rational addiction, there is evidence that they are not fully forward-looking.

These findings have important implications for design of health insurance policies, especially in light of the Affordable Care Act (ACA). The majority of the plans offered on the health insurance exchanges, created by the ACA, have a deductible, and for many policies the deductibles are quite large. Moreover, a growing number of US employers offer plans with at least some deductible (Kaiser Family Foundation, 2012). Thus, going forward, demand for medical care cannot be evaluated with a single price of care, and both research and policy-making must account for the nonlinearities created by deductibles. This research highlights that individuals understand and respond to the incentives created by the presence of a deductible in the insurance policy. This is reassuring and confirms that cost-sharing provides the intended incentives.

What is not as clear is whether the size of the deductible matters when individuals make decisions about medical care. The findings in this study show that reaching the deductible, whether via one large health event or a series of smaller health visits that add up to the deductible amount, plays an important role in decision-making by lowering end-of-year price, which consumers respond to. However, the results are not as clear on whether the size of deductibles makes a significant impact. Admittedly, the deductibles in the sample are smaller compared to what is reported on average (see Kaiser Family Foundation, 2012). Since reaching the deductible has a robust strong impact on spending, the inclusion of deductibles into policies may indeed play its intended role in preventing overconsumption but not deterring those with "needed" spending from using medical care.

A notable implication for policy is that deductibles of reasonable size (i.e. compared to incomes) may not be a significant determinant of demand for those who know that they will reach and exceed the deductible, as in the case of pregnant women who give birth. If a large health event is expected, and hence a lower end-of-year price, consumers may not face any deterrents for using medical care. If the main concern is circumventing excessive spending, coinsurance may be a better tool to deter high spenders. In the sample used in the study the average coinsurance rate is about 0.13. Kaiser Family Foundation (2012) reports that the average coinsurance for covered workers is around 0.18-0.19. This is relatively low, and may actually present an incentive to get to that portion of the budget constraint (see Figure 1). Raising the coinsurance rate to a more substantial rate could affect the end-of-year price and hence demand for care. However, to ensure that care is still affordable and accessible, the coinsurance rate should not be overly high, considering that individuals who would be in the coinsurance rate of the budget constraint may be in this region due to a large health event and may need to continue a course of care. Hence, there is a delicate balance between the size of the deductible and the coinsurance rate. Some earlier studies (see Feldstein, 1973; Manning and Marquis, 1996) have examined the effect of deductible-coinsurance combinations on the demand, but in a static spot price setting. This balance has to be placed and investigated in the correct framework: a consumer who faces different prices throughout the year in the context of uncertainty.

While this study undoubtedly contributes to the scarce literature on nonlinear pricing and the demand for medical care, the findings are based on a very specific subsample of the employer-insured population. On the one hand, normal pregnancy is a homogenous health condition, with a predictable large health event for those who give birth. This is why this sub-population was chosen here, especially as one of the objectives is to examine forward-looking behavior. However, it is also a particular health condition. First, pregnant women are not advised medically to purchase many discretionary non-pregnancy related procedures, which makes their price responsiveness less generalizable. Even though the selection of non-pregnancy expenditures and visits was done conservatively to avoid mixing in pregnancy-related care, due to the nature of the claims data and medical billing it is not unreasonable to assume that some of the pregnancy-related procedures were in fact not related to pregnancy, and vice-versa. This could add some measurement error to the results.

Finally, labor itself, despite being a predictable large health event, creates a particular health context that is affected by other factors and in turn influences the demand through price and other channels. Hence, the demand in the labor group is more sensitive, to price and other factors, than the demand among non-labor pregnant women. For example, the results in Chapter 6 show evidence of myopia only among the women in the labor group, through the month of labor effect.[1] Thus, other factors in addition to price, such as the change in the availability of a woman for doctor visits, may affect demand for care in this context. In fact, these women may be forward-looking not just in regards to price, but other variables.

To test the overall robustness of the findings about forward-looking behavior and to make these results more generalizable, a similar empirical analysis should be done using other health conditions, both predictable, such as management of chronic diseases, and less certain, such as an onset of a chronic condition or an injury due to an accident. Applying these research questions and methods to other health scenarios across the predictability spectrum would be a useful extension of this study and will in a way calibrate the findings on anticipatory effects, which are in large part driven by the ability to predict the end-of-year price. Moreover, it would be useful to compare the price responsiveness across health conditions to better inform the design of health insurance

[1] Even though myopia was not examined specifically among the non-labor women, the results in Chapter 7, which includes the non-labor women who experienced a price change through reaching the deductible, show that it is likely that the women who did not give birth but reached the deductible are foward-looking.

policies.

References

[1] Aron-Dine, A., Einav, L., Finkelstein, A. and M. Cullen (2012). Moral Hazard in Health Insurance: How Important Is Forward Looking Behavior? National Bureau of Economic Research, NBER Working Paper 17802 http://www.nber.org/papers/w17802.

[2] Aron-Dine, A., Liran E,, and A. Finkelstein. 2013. The RAND Health Insurance Experiment, Three Decades Later. *Journal of Economic Perspectives*, 27(1): 197-222.

[3] Arrow, K. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, 53(5): 941-973.

[4] Basu, A. (2005). Extended generalized linear model: Simultaneous estimation of flexible link and variance functions. *The Stata Journal* 5 (4): 501-516.

[5] Basu, A. and P. J. Rathouz. (2005). Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. *Biostatistics* 6(1): 93?109.

[6] Belotti, F. and P. Deb 2012. TPM: Stata Module to Estimate Two-Part Cross-Sectional Models. *The Stata Journal* 20: 1-13.

[7] Blundell, R. and T. MaCurdy. (1999). "Labor Supply: A Review of Alternative Approaches" in Ashenfelter, O. and D. Card (ads). *Handbook of Labor Economics.* Oxford: Elsevier North Holland.

[8] Borenstein, S. (2009). To what electricity price do consumers respond? Residential demand elasticity under increasing-block pricing. Mimeo, UC Berkeley.

[9] Burtless, G. and J. Hausman. (1978). The effect of taxation on labor supply: Evaluating The Gary Negative Income Tax Experiment. *Journal of Political Economy* 86 (6): 1103-1130.

[10] Cameron, C. and P. Trivedi. (2009). *Microeconometrics Using Stata*. College Station, Texas: Stata Press.

[11] Chevalier, J. and A. Goolsbee (2009). Are durable goods consumers forward-looking? Evidence from college textbooks. *Quarterly Journal of Economics* 124(4): 1853-1884.

[12] Deb, P., Manning, W. and E. Norton. (2013). Modeling Health Care Costs and Counts. Presented at iHEA World Congress, Sydney, Australia.

[13] Doshi, J., Glick, H. and D. Polsky. 2006. Analyses of Cost Data in Economic Evaluations Conducted Alongside Randomized Controlled Trials. *Value in Health* 9(5): 334-340.

[14] Duan, N. (1983). Smearing Estimate: A Nonparametric Retransformation Method. *Journal of the American Statistical Association* 78 (383): 605-610.

[15] Ellis, R. (1986). Rational behavior in the presence of coverage ceilings and deductibles. *The RAND Journal of Economics* 17(2): 158-175.

[16] Ellis, R. and T. McGuire. (1986). Cost sharing and patterns of mental health care utilization. *Journal of Human Resources* 21(3): 359-379.

[17] Feldstein, M. (1973). The welfare cost of excess health insurance. *Journal of Political Economy* 81(2): 251-280.

[18] Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, 80(2), 223-255.

[19] Gruber, J. and B. Köszegi. (2001). Is addiction "rational"? Theory and evidence. *Quarterly Journal of Economics* 114 (4): 1261-1303.

[20] Hausman, J. (1979). Individual discount rates and the purchase and utilization of energy using durables. *Bell Journal of Economics* 10: 33-54.

[21] Hausman, J. (1985). The econometrics of nonlinear budget sets. *Econometrica* 53 (6): 1255-1281.

[22] Hausman, J., Kinnucan, M., and D. McFadden. (1979). A two-level electricity demand model: Evaluation of the Connecticut time-of-day pricing test. *Journal of Econometrics* 8: 263-289.

[23] Hausman, J. and J. Trimble. (1984). Appliance purchase and usage adaptation to a permanent time of day electricity rate schedule. *Journal of Econometrics* 26: 115-139. Houthakker, H. (1951). Electricity tariffs in theory and practice. *Economic Journal* 61: 1-25.

[24] Jones, A. 2010. Models for health care. Heath Econometrics and Data Group (HEDG) Working Paper 10/1. The University of York, York, UK.

[25] Jung, K., Feldman, R., and M. McBean. 2014a. Demand for prescription drugs under non-linear pricing in Medicare Part D. *International Journal of Health Care Finance and Economics* 14 (1): 19-40.

[26] Jung, K., Feldman, R. and M. McBean. 2014b. Non-linear Pricing in Drug Benefits and Medication Use: The Case of Statin Compliance in Medicare Part D. *Health Services Research* 49 (3): 910-928.

[27] Kaiser Family Foundation and Health Research and Educational Trust. (2012). Employer Health Benefits: 2012 Annual Survey. Henry J. Kaiser Family Foundation, Menlo Park, CA and Health Research and Educational Trust, Chicago, IL.

[28] Keeler, E., Newhouse, J., and C. Phelps. (1977). Deductibles and the demand for medical care services: The theory of a consumer facing a variable price schedule under uncertainty. *Econometrica* 45(3): 641-656.

[29] Keeler, E. and J. Rolph. (1988). The demand for episodes of treatment in the Health Insurance Experiment. *Journal of Health Economics* 7: 337-367.

[30] Kowalski, A. (2009). Censored quantile instrumental variable estimates of the price elasticity of expenditure on medical care. NBER Working Paper 15085.

[31] Kowalski, A. (2012). Estimating the tradeoff between risk protection and moral hazard with a nonlinear budget set model of health insurance. National Bureau of Economic Research, NBER Working Paper 17802 http://www.nber.org/papers/w18108.

[32] Kuklina, E., Whiteman, M., Hillis, S., Jamieson, D., Meikle, S., Posner, S., and P. Marchbanks. (2008). An enhanced method for identifying obstetric deliveries: Implications for estimating maternal morbidity. *Maternal and Child Health Journal* 12: 469-477.

[33] Manning, W. (1998). The logged dependent variable, heteroskedasticity, and the retransformation problem. *Journal of Health Economics* 17: 283-295.

[34] Manning, W. and J. Mullahy. Estimation log models: to transform or not to transform? *Journal of Health Economics* 20: 461-494.

[35] Manning, W. and S. Marquis. (1996). Health insurance: The tradeoff between risk pooling and moral hazard. *Journal of Health Economics* 15(5): 609-640.

[36] Manning, W., Newhouse, J., Duan, N., Keeler, E., Leibowitz, A., and S. Marquis (1987). Health insurance and the demand for medical care: Evidence from a randomized experiment. *American Economic Review* 77(3): 251-277.

[37] Manning, W., Wells, K., Duan, N., Newhouse, J. and J. Ware. (1986). How cost sharing affects the use of ambulatory mental health services. *JAMA* 256 (14): 1930-1934.

[38] MarketScan Database, Thompson Medstat. (2004). Ann Arbor, Michigan.

[39] Marsh, C. (2013). Estimating demand elasticities using nonlinear pricing. University of Georgia: Unpublished paper.

[40] McCullagh, P. and J. Nelder. 1989 *Generalized Linear Models, Second Edition.* CRC Press, Boca Raton, Florida.

[41] Mullahy, J. (1997). Instrumental variable estimation of count data models: Applications to models of cigarette smoking behavior. *Review of Economics and Statistics* 79: 586-593.

[42] Mullahy, J. (1998). Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *Journal of Health Economics* 17: 247-281.

[43] Nichols, A. (2007). ivpois: Stata module for IV/GMM Poisson regression. http://ideas.repec.org/c/boc/bocode/s456890.html

[44] Pauly, M. (1968). The economics of moral hazard: Comment. *American Economic Review* 58: 531-537.

[45] Pauly, M. and F. Blavin. (2008). Moral hazard in insurance, value-based cost sharing, and the benefits of blissful ignorance. *Journal of Health Economics* 27(6): 1407-1417.

[46] Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy* 2: 180-212.

[47] Seoud, M., Nassar, A., Usta I., Melhem Z., Kazma A., and A. Khalil. 2002. Impact of advanced maternal age on pregnancy outcome. *American Journal of Perinatology* 19 (1): 1-8.

[48] Strotz, R. (1956). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies* 23 (3): 165-180.

[49] Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data.* The MIT Press: Cambridge, MA.

[50] Van Vliet, R. (2001). Effect of price and deductibles on medical care demand, estimated from survey data. *Applied Economics* 33: 1515-1524.

[51] Zeckhauser, R. (1970). Medical insurance: A case study of the tradeoffs between risk spreading and appropriate incentives. *Journal of Economic Theory* 2(1): 10-26.